The Finite State MAC with Cooperative Encoders and Delayed CSI

Ziv Goldfeld, Haim H. Permuter and Benjamin M. Zaidel

Abstract

In this paper, we consider the finite-state multiple access channel (MAC) with partially cooperative encoders and delayed channel state information (CSI). Partial cooperation here is in the sense that the encoders communicate with each other through finite-capacity links. The channel states are assumed to be governed by a Markov process. Full CSI is assumed at the receiver, while only delayed CSI is available at the transmitters. The capacity region of this model is derived by first solving the case of the finite-state MAC with common message. Achievability for the common message case is established using rate splitting, multiplexing and simultaneous decoding. Simultaneous decoding is crucial here since it circumvents the need to rely on the capacity region's corner points, which becomes cumbersome as the number of messages to be sent grows. The common message result is then used to derive the capacity region for the case with partially cooperating encoders. Next, we apply this general result to the special case of the Gaussian vector MAC with diagonal channel transfer matrices, which is suitable for modeling, e.g., orthogonal frequency division multiplexing (OFDM)-based communication systems. The capacity region of the Gaussian channel is presented in terms of a convex optimization problem, which can be solved efficiently using numerical tools. The region is derived by first presenting an outer bound on the general capacity region, and then suggesting a specific input distribution that achieves this bound. Finally, numerical results are provided that give valuable insights into the practical implications of optimally using conferencing in order to maximize the transmission rates.

Index Terms

Capacity region, Common message, Convex optimization, Cooperative encoders, Delayed CSI, Diagonal vector Gaussian Multiple-access channel, Finite-state channel, Fourier-Motzking elimination, Multiple-access channel, Simultaneous decoding.

I. INTRODUCTION

Time-varying channels and their research have been drawing increasing attention over the past few years. This is due to the fact that these channels successfully model wireless communication systems, which constitute the most prevalent form of communication today. In a wireless setting, the user's motion and the changes in the environment, as well as interference, may lead to temporal changes in the channel quality. The time-varying channel

Ziv Goldfeld and Haim Permuter are with the department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel. Author emails: zgzg1984@gmail.comm, haimp@bgu.ac.il, benjamin.zaidel@gmail.com.

This paper was presented in part at ISIT 2012, Cambridge, MA, USA.

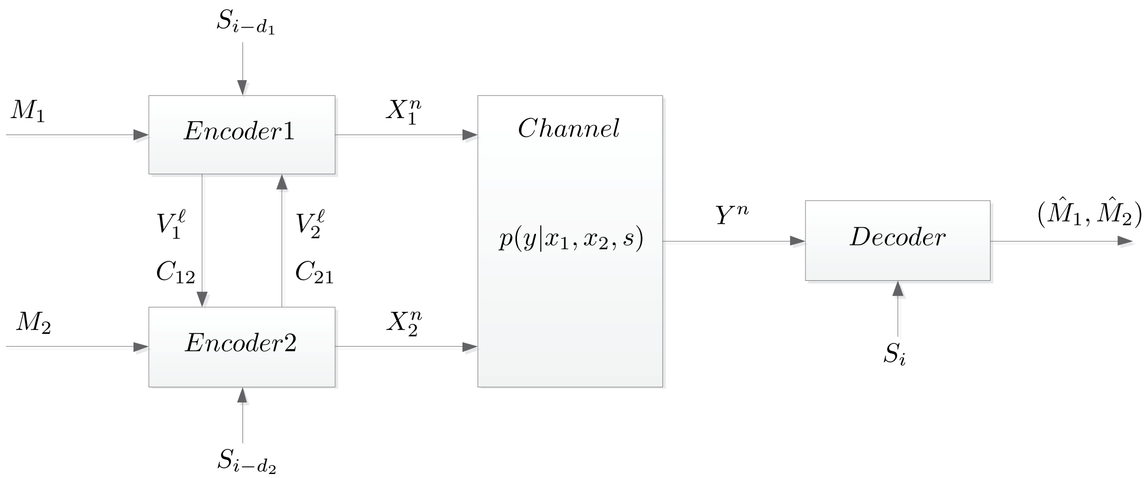


Fig. 1: FSM-MAC with partially cooperative encoders, CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2.

characteristics give rise to the need for channel state information (CSI) estimation. For example, the Long Term Evolution (LTE) cellular communication standard uses pilot signals that are transmitted at pre-scheduled time intervals and frequency slots in order to estimate the channel's state [1]. These estimations are performed at the receiver and then commonly fed back to the transmitter. Frequently, the feedback is not instantaneous, which results in the fact that the transmitter has access to delayed CSI. In addition, it seems that future wireless communication is heading towards user cooperation in order to achieve enhanced performance. In view of the above, we explore in this paper the impact of both cooperation and the availability of CSI. We focus on a setting of a finite state Markov (two-user) multiple access channel (FSM-MAC), with partially cooperative encoders and delayed CSI, as illustrated in Fig. 1 and explained in the following.

In the communication scenario under discussion, each of the two encoders wishes to send an independent private message through a time-varying MAC to the decoder. Delayed CSI is assumed to be available at the encoders, while full delayless CSI is assumed at the decoder. Different users may be subject to different CSI delays. It is further assumed that prior to each transmission block, the two encoders are allowed to hold a conference. More specifically, it is assumed that the encoders can communicate with each other over noise-free communication links of given capacities. We restrict the discussion to the case in which the conference held between the encoders is independent of the CSI.

The non-state-dependent MAC with partially cooperative encoders was first introduced by Willems [2], who also derived the capacity region for the discrete memoryless setting. Special cases of this channel model include the case in which the encoders are ignorant of each other's messages (i.e., the capacities of the communication links between them are both zero), and the case where the encoders fully cooperate (i.e., the capacities of the communication links are infinite). The first setting, where no conference is held, corresponds to the classical MAC, for which the capacity region was determined by Ahlswede [3] and Liao [4]. In contrast, in the second setting, where total

cooperation is available, the encoders can act as one by fully sharing their private messages via the conference. The capacity region for this case is the part of the first quadrant below the so-called total cooperation line. This triangle-shaped region always contains the capacity region for the classical MAC.

In his proof of achievability for the conferencing MAC in [2], Willems produced a coding scheme based on the capacity region for the MAC with common message. This problem was defined and solved by Slepian and Wolf [5]. Willems showed that in order to achieve the capacity region, the encoders should use the cooperation link to share parts of their private messages and then use a coding scheme for the ordinary MAC with common message. Although Willems's model allows interactive communication between the encoders, it was shown both in [2], and later in [6], that the optimality is achieved even in a single round of communication between the encoders (referred to as a "pair of simultaneous monologues" in [2]).

Additional multiuser settings that involve cooperation between users through communication links of finite capacities have been extensively treated in the literature. The study on such channels includes works on the MAC [7], [8], interference channel [9]–[15], broadcast channel [16], relay channel [17], [18], and cellular networks [19]. A comprehensive survey of cooperation and its role in communication is given in [20]. It is important to note, however, that in all of the above settings the channel was not assumed to be time-varying.

Naturally, multiuser settings combining time-varying channels and user cooperation were the next juncture in research. A Gaussian fading MAC with cooperating encoders that have access to delayless CSI was considered in [21], [22]. As in our case, these works assumed that the cooperation is allowed only before the CSI becomes available at the encoders. A different approach, in which the cooperation occurs after the state information becomes available, was treated in [23]. In this work, a MAC with perfect non-causal CSI available at the encoders was considered. The coding scheme presented in [23] used the conferencing in order to share parts of the messages as well as the state information regarding the channel's variation in time.

The notion of modeling time-varying channels as state-dependent channels dates back to Shannon [24]. In that work, Shannon introduced and characterized the capacity of the state-dependent, memoryless point-to-point channel with independent and identically distributed (i.i.d.) states available causally at the encoder. Gel'fand and Pinsker [25] and, later, Heegard and El Gamal [26], studied the case where the encoder observes the channel states non-causally. They derived a single letter formula for the capacity using a binning coding scheme. In [27], Goldsmith and Varaiya considered a fading channel with perfect CSI at both transmitter and receiver. They showed that given the instantaneous and perfect state information the transmitter can adapt the data rates to each of the channel's states, thus maximizing the average transmission rate.

The impracticality of perfect CSI steered the research of state-dependent channels to consider models involving imperfect CSI. At first, different cases of imperfect CSI of an i.i.d. state sequence were treated. Caire and Shamai [28] considered a state-dependent model in which the CSI at the transmitter was assumed to be a deterministic function of the CSI at the receiver. They have managed to show that the optimal coding scheme is particularly simple. Namely, it was shown that optimal codes can be constructed directly over the input alphabet, while in general, coding over an expanded alphabet is required. In [29], Lapidoth and Steinberg provided an inner bound

for the capacity region of the MAC with strictly causal CSI at the encoders. As opposed to the point-to-point case, where strictly causal CSI regarding an i.i.d. state sequence does not increase capacity, the MAC's capacity region is strictly increased as a result of it. Li *et al*. presented an improved inner bound for the same setting in [30]. The innovative idea of an information theoretical model involving delayed CSI, where the states are not i.i.d., was first introduced by Viswanathan who presented and solved the FSM point-to-point channel with delayed CSI [31]. In a similar manner to Viswanathan, we model temporal variations by means of a FSM channel [32], [33]. The channel state is determined on a per symbol basis and governed by the underlying FSM process.

As research of FSM channels with delayed CSI continued to gain momentum, an important extension of this novel idea to the multiuser case was introduced. In [34], Basher *et al*. considered the FSM-MAC with delayed CSI and ignorant encoders, i.e., where no conference is held [34] (see also [35] for a related source coding analysis). In the proof of the capacity region for this model, achievability was established by employing a coding scheme based on successive decoding. Successive decoding was used in order to demonstrate that the two corner points of the capacity region are achievable. The whole capacity region is then achievable via time-sharing.

In our setting, where conferencing takes place, a different approach is needed. Since the achievability for the conferencing model is based on the common message coding scheme [2], we start by solving the FSM-MAC with common message and the same CSI properties as in [34], which remained an open problem until the current paper. Next, using the achievable scheme of the common message setting, the achievability of the conferencing region is established. However, providing an achievable coding scheme for the common message setting based on achieving the region's corner points has turned out to be an awkward task. This is due to the large number of corner points which are induced by the presence of an additional transmission rate, namely the rate of the common message. Therefore, we present a more general scheme that achieves every possible point in the region, rather than just the corner points. We use rate-splitting and multiplexing-coding in the encoding stage. Since these ideas were also used in [34], the structure of the encoders in both schemes is similar. The *decoding* process, on the other hand, is utterly different because we consider a scheme by which the decoder decodes all the messages *simultaneously*. Simultaneous decoding gives rise to great difficulties when analyzing the probability of error, yielding a very large number of inequalities for the partial rates. Fortunately, using the Fourier-Motzkin elimination, the partial rates can be eliminated, and the rate constraints can be expressed via a small number of inequalities, which specify the capacity region. The above simultaneous decoding scheme for the MAC with delayed CSI is one of the most significant contributions of our paper. Not only does it generalize the coding scheme presented in [34] and can be used for the ignorant encoders setting, but also it sets the ground for constructing coding schemes for a general number of users. This construction can be done by a direct and trivial extension of the scheme we present here.

Based on these general results, we then derive the capacity region for the special case of a vector Gaussian FSM-MAC with diagonal channel transfer matrices. This channel model can be used to represent an orthogonal frequency-division multiplexing (OFDM)-based communication system, employing single receive and transmit antennas. The diagonal entries of the channel matrices represent the orthogonal sub-channels used by the OFDM scheme.

In order to derive the capacity region for the above channel, we use a multivariate extension of a novel tool first

derived in [36]. Using this tool, we demonstrate that Gaussian multivariate distributions maximize certain mutual information expressions under a Markovity constraint. The scalar version of this tool was employed by Lapidoth *et al*. [37] to provide an outer bound for the capacity region of the scalar Gaussian non-state-dependent MAC with conferencing encoders. Wigger and Kremer also used this tool in their solution for the capacity region of the three-users non-state-dependent MIMO MAC with conferencing [38]. The reason why such a tool is needed originates in the fact that the input distribution of the conferencing channel must admit a certain Markovity constraint. In cases where no Markov relation is to be satisfied, the traditional approach for proving the optimality of Gaussian multivariate distributions is by employing either the Vector Max-Entropy Theorem (a direct extension of [39, Theorem 12.1.1]) or a conditional version of it. Here, however, this approach fails since replacing a non-Gaussian vector satisfying the Markovity condition by a Gaussian vector of the same covariance matrix may result in a Gaussian vector that violates the Markovity condition. In order to overcome this issue we use a sufficient and necessary condition on the (auto- and cross-) covariance matrices of the involved Gaussian random vectors (RVs) in order for them to admit a Markov relation [40, Section 2, Theorem 1].

We note that although Gaussian input vectors are shown to be optimal in this setting, the original form of the capacity region involves a non-convex optimization problem. To alleviate this difficulty, new variables are introduced to convert the optimization problem into a convex one, which can then be solved using numerical tools such as CVX [41]. The capacity region for the corresponding scalar Gaussian channel can be immediately derived from the result for the vector channel setting, and serves as an extension of the result in [37] to the state-dependent case. The capacity region of the vector Gaussian FSM-MAC with common message and the same CSI properties is also easily derivable from the result for the conferencing channel, by exploiting the strong correspondence between the two models and using a simple analogy.

We conclude this paper with a specific example, namely a scalar AWGN channel with two possible states (‘Good’ and ‘Bad’), in order to gain some insights into the practical implications of the results. Numerical results are included to demonstrate the impact of different channel parameters on the capacity region and the optimal input distribution. The interactions between the different parameters are interpreted in a manner that produces valuable insights.

The remainder of the paper is organized as follows. In Section II we describe the two communication models of interest - the FSM-MAC with common message and delayed CSI, as well as the FSM-MAC with partially cooperative encoders and delayed CSI. In Sections III and IV, we state the capacity results for the common message and conferencing models, respectively. Each result is followed by its proof. Next, in Section V, the vector Gaussian FSM-MAC with diagonal channel transfer matrices is defined and the maximization problem defining its capacity region is derived. The regions for the corresponding common message model and the scalar setting are given as special cases. The two-state Gaussian example is discussed in this section as well. Finally, Section VI summarizes the main achievements and insights presented in this work along with some possible future directions and extensions.

II. CHANNEL MODELS AND NOTATIONS

In this paper, we investigate the capacity region of the FSM-MAC with partially cooperative encoders, full CSI at the decoder (receiver) and delayed CSI at the encoders (transmitters), as illustrated in Fig. 1. In order to do so, we first consider a different setting, which is the FSM-MAC with a common message and the same CSI properties, see Fig. 2. The derivation of the capacity region for the common message setting forms the basis for the achievability proof for the conferencing case. Since most definitions for both channels follow similar lines, we start by defining the common message setting and then extend the description for the setting of partially cooperative encoders.

Throughout this work we use the following notations. Scalars are denoted by lower case letters whereas random variables are denoted by upper case letters. We denote deterministic column vectors by boldface lower case letter, namely $\mathbf{x} = (x_1, \ldots, x_N)^T$ (the vector's dimensions will be stated explicitly), while random column vectors are denoted by boldface upper case letters as $\mathbf{X}$. Matrices are denoted by non-italic upper case letters, that is X, and finally, we use X^n to denote the sequence $\{X_1, \ldots, X_n\}$.

A. FSM-MAC with Common Message and Delayed CSI

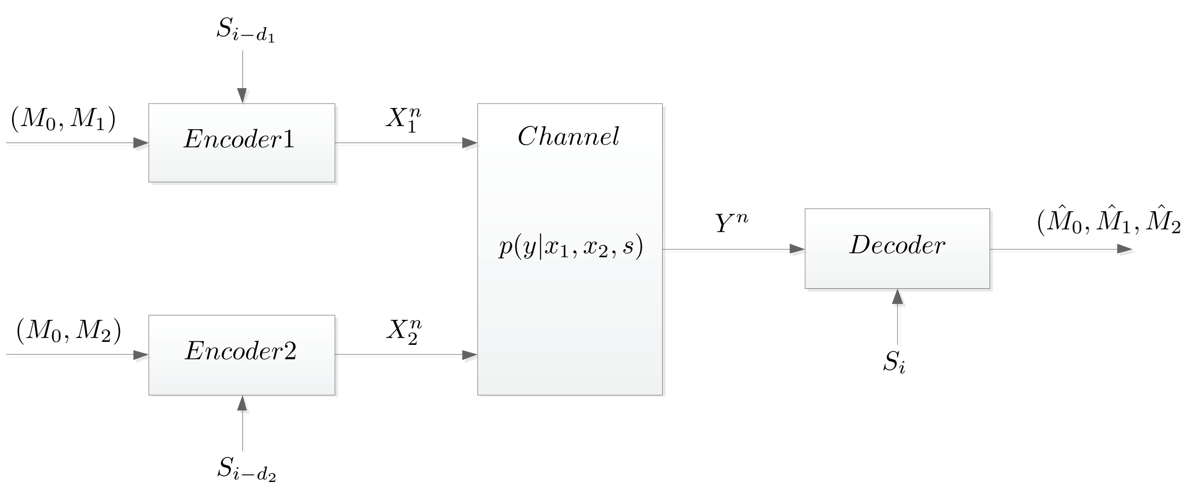


Fig. 2: FSM-MAC with a common message, full CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2.

We consider the communication system of FSM-MAC with a common message, CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2, as illustrated in Fig. 2. The MAC setting consists of two senders and one receiver. Each sender $j \in \{1, 2\}$ chooses a pair of indices, (m_0, m_j), uniformly from the set $\{1, ..., 2^{nR_0}\} \times \{1, ..., 2^{nR_j}\}$, where m_0 denotes the common message and m_j, $j \in \{1, 2\}$, denotes the private message of the corresponding sender. The choices of m_0, m_1 and m_2 are independent.

The input to the channel from encoder $j \in \{1, 2\}$ is denoted by $\{X_{j,1}, X_{j,2}, \ldots, X_{j,n}\}$, and the output of the channel is denoted by $\{Y_1, Y_2, \ldots, Y_n\}$. Using the notations stated at the beginning of this section we simply denote the channel's inputs and output by X_j^n, $j \in \{1, 2\}$ and Y^n, respectively.

The FSM channel is assumed to be, at each time instance, in one of a finite number of states $\mathcal{S}=\{s_1,s_2,...,s_k\}$. In each state, the channel is a discrete memoryless channel (DMC), with input alphabets $\mathcal{X}_1,\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. Let the random variables S_i and S_{i-d_j}, $j\in\{1,2\}$, denote the channel state at times i and $i-d_j$, respectively. Similarly, we denote by $X_{1,i},X_{2,i}$ and Y_i the inputs and the output of the channel at time i. The channel transition probability distribution at time i depends on the state S_i, and the inputs $X_{1,i},X_{2,i}$ at time i, and is given by $P(y_i|x_{1,i},x_{2,i},s_i)$. The channel output at any time i is assumed to depend only on the channel inputs and state at time i. Hence,

$$P(y_i|x_1^i,x_2^i,s^i)=P(y_i|x_{1,i},x_{2,i},s_i). \tag{1}$$

The state process, $\{S_i\}_{i=1}^n$, is assumed to be an irreducible, aperiodic, finite-state homogeneous Markov chain and is therefore ergodic. The state process is independent of the channel inputs and output when conditioned on the previous states, i.e.,

$$P(s_i|s^{i-1},x_1^{i-1},x_2^{i-1},y^{i-1})=P(s_i|s_{i-1}). \tag{2}$$

Furthermore, we assume that the state process is independent of the messages M_0, M_1 and M_2, i.e.,

$$P(s^n,m_0,m_1,m_2)=\prod_{i=1}^n P(s_i|s_{i-1})P(m_0)P(m_1)P(m_2). \tag{3}$$

Now, let K be the one step state-transition probability matrix of the Markov process and let π be its steady state probability distribution. The joint distribution of (S_i,S_{i-d}) is stationary and is given by

$$\pi_d(S_i=s_l,S_{i-d}=s_j)=\pi(s_j)\mathrm{K}^d(s_l,s_j), \tag{4}$$

where $\mathrm{K}^d(s_l,s_j)$ is the (l,j)-th element of the d-step transition probability matrix, K^d, of the Markov state process. Without loss of generality, we assume henceforth that $d_1\geq d_2$. Furthermore, to simplify the notation, we define the joint distribution of the variables $(S,\tilde{S}_1,\tilde{S}_2)$ as the joint distribution of the random variables (RVs) $(S_i,S_{i-d_1},S_{i-d_2})$, i.e.,

$$P_{S\widetilde{S}_1\widetilde{S}_2}(s_l,s_j,s_v)=\pi(s_j)\mathrm{K}^{d_1-d_2}(s_v,s_j)\mathrm{K}^{d_2}(s_l,s_v), \tag{5}$$

where $(s_j,s_l,s_v)\in\mathcal{S}^3$. A $(n,2^{nR_0},2^{nR_1},2^{nR_2},d_1,d_2)$ code for the FSM-MAC with CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2 consists of:

1) Three sets of integers $\mathcal{M}_0=\{1,2,...,2^{nR_0}\}$, $\mathcal{M}_1=\{1,2,...,2^{nR_1}\}$ and $\mathcal{M}_2=\{1,2,...,2^{nR_2}\}$, referred to as the *message sets*.
2) Two encoding functions f_j, $j\in\{1,2\}$. Each function f_j is defined by means of a sequence of functions $f_{j,i}$ (where i denotes the time instance) that depend only on the pair of messages (M_0,M_j), and the channel

states up to time $i-d_j$. The output of Encoder j at time i is given by

$$X_{j,i} = \begin{cases} f_{j,i}(M_0, M_j), & 1 \le i \le d_j \\ f_{j,i}(M_0, M_j, S^{i-d_j}), & d_j + 1 \le i \le n \end{cases}. \tag{6}$$

3) A decoding function:

$$\psi : \mathcal{Y}^n \times \mathcal{S}^n \to \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \ . \tag{7}$$

We define the average probability of error for the $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code as

$$P_e^{(n)} = \frac{1}{2^{n(R_0+R_1+R_2)}} \sum_{m_0,m_1,m_2} \sum_{s^n} P_{S^n}(s^n) \mathbb{P}\left\{\psi(y^n, s^n) \neq (m_0, m_1, m_2) | (m_0, m_1, m_2) \text{ was sent}\right\}, \tag{8}$$

where $\mathbb{P}\{\mathcal{A}\}$ denotes the probability of the event $\mathcal{A}$.

We use standard definitions [39] of achievability and of the capacity region, namely, a rate triplet (R_0, R_1, R_2) is *achievable* for the FSM-MAC if there exists a sequence of $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes with $P_e^{(n)} \to 0$ as n goes to infinity. The *capacity region* is the closure of the set of achievable rates (R_0, R_1, R_2).

B. FSM-MAC with Partially Cooperative Encoders and Delayed CSI

We now define the FSM-MAC with CSI at the decoder, delayed CSI at the encoders with delays d_1 and d_2, and partially cooperative encoders with cooperation (or conferencing) links of capacities C_{12} and C_{21}, as illustrated in Fig. 1.

The channel definition relies on Subsection II-A, while taking the common message set to be $\mathcal{M}_0 = \emptyset$. Here, however, conferencing between the encoders is introduced. The conference is assumed to take place prior to the transmission of a codeword through the channel and consists of ℓ consecutive pairs of communications, simultaneously transmitted by the encoders. Each communication depends on the message to be transmitted by the sending encoder and previously *received* communications from the other encoder. We denote the communications transmitted from encoder $j \in \{1, 2\}$ to the other encoder by V_j^ℓ. Note that here the state process is also assumed to be independent of the conference communications, i.e.,

$$P(s^n, v_1^\ell, v_2^\ell) = P(s^n) P(v_1^\ell, v_2^\ell) = \prod_{i=1}^{n} P(s_i | s_{i-1}) P(v_1^\ell, v_2^\ell). \tag{9}$$

A $(n, \ell, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code for the FSM-MAC with CSI at the decoder, delayed CSI at the encoders with delays d_1 and d_2, and conferencing links with capacities C_{12} and C_{21} consists of:

1) Two sets of integers $\mathcal{M}_1 = \{1, 2, ..., 2^{nR_1}\}$ and $\mathcal{M}_2 = \{1, 2, ..., 2^{nR_2}\}$, referred to as the *message sets*.
2) Two encoders, where each encoder is completely described by an encoding function, f_j, and a set of ℓ ($\ell \ge 1$) communication functions, $\{h_{j,1}, h_{j,2}, \ldots, h_{j,\ell}\}$, $j \in \{1, 2\}$ (similar definitions were also used in [2]).
3) The encoding function, f_j, maps the message M_j, $j \in \{1, 2\}$, and what was learned from the conference with

the other encoder into channel codewords of length n. Each function f_j is defined by means of a sequence of functions $f_{j,i}$ that depend only on the message M_j, the received communications from the other encoder in the conferencing stage and the channel states up to time $i-d_j$. We emphasize that since the encoding occurs only after the conferencing has finished, each $f_{j,i}$ depends on all received communications.

4) Each communication function $h_{1,i}$ (respectively, $h_{2,i}$), $i \in \{1,2,\ldots,\ell\}$, maps the message M_1 (respectively, M_2) and the sequence of previously received communications from the other encoder V_2^{i-1} (respectively, V_1^{i-1}), onto the i-th communication $V_{1,i}$ (respectively, $V_{2,i}$). More specifically, the communications are defined as:

$$V_{1,i} = h_{1,i}(M_1, V_2^{i-1}) \; ; \; V_{2,i} = h_{2,i}(M_1, V_1^{i-1}). \tag{10}$$

5) The encoding function for Encoder 1 satisfies

$$X_{1,i} = \begin{cases} f_{1,i}(M_1, V_2^{\ell}), & 1 \leq i \leq d_1 \\ f_{1,i}(M_1, V_2^{\ell}, S^{i-d_1}), & d_1 + 1 \leq i \leq n \end{cases}, \tag{11}$$

and the encoding function for Encoder 2 is defined in an analogous manner (using the private message M_2, the communications V_1^{ℓ} and the delay d_2).

6) The random variable $V_{j,i}$, for $j \in \{1,2\}$ and $i \in \{1,2,\ldots,\ell\}$ ranges over the finite alphabet $\mathcal{V}_{j,i}$. In the case of partially cooperating encoders, the amount of information exchanged during the conference is bounded by the finite communication link capacities C_{12} and C_{21}. A conference is (C_{12}, C_{21})-permissible if the sets of communication functions are such that [2]:

$$\sum_{i=1}^{\ell} \log |\mathcal{V}_{1,i}| \leq nC_{12} \; ; \; \sum_{i=1}^{\ell} \log |\mathcal{V}_{2,i}| \leq nC_{21}. \tag{12}$$

7) A decoding function:

$$\psi : \mathcal{Y}^n \times \mathcal{S}^n \to \mathcal{M}_1 \times \mathcal{M}_2. \tag{13}$$

The average probability of error for the $(n, \ell, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code is

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{m_1,m_2} \sum_{s^n} P_{S^n}(s^n) \mathbb{P}\{\psi(y^n, s^n) \neq (m_1, m_2) | (m_1, m_2) \text{ was sent}\}. \tag{14}$$

The achievable rates and the capacity region are defined in an analogous manner to Section II-A.

III. THE CAPACITY REGION OF THE FSM-MAC WITH A COMMON MESSAGE AND DELAYED TRANSMITTER CSI

In this section we state the capacity region of the FSM-MAC with common message and delayed transmitter CSI, followed by it's proof. Without loss of generality, we assume that $d_1 \geq d_2$.

Theorem 1 *The capacity region of the FSM-MAC with a common message, CSI at the decoder and asymmetrically delayed CSI at the encoders with delays d_1 and d_2, as defined in Section II-A and shown in Fig. 2, is given by:*

$$\mathcal{C}_{CM} = \bigcup_{P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)} \left\{ \begin{array}{rcl} R_1 & \leq & I(X_1;Y|X_2,U,S,\tilde{S}_1,\tilde{S}_2), \\ R_2 & \leq & I(X_2;Y|X_1,U,S,\tilde{S}_1,\tilde{S}_2), \\ R_1+R_2 & \leq & I(X_1,X_2;Y|U,S,\tilde{S}_1,\tilde{S}_2), \\ R_0+R_1+R_2 & \leq & I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2). \end{array} \right\}, \tag{15}$$

where U is a RV with bounded cardinality and where the joint distribution of $(S,\widetilde{S}_1,\widetilde{S}_2)$ is specified in (5).

A. Converse

Given an achievable rate triplet (R_0,R_1,R_2) we need to show that there exists a joint distribution of the form $P(s,\tilde{s}_1,\tilde{s}_2)P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)P(y|x_1,x_2,s)$ such that the inequalities in (15) are satisfied. Since (R_0,R_1,R_2) is an achievable rate triplet, there exists an $(n,2^{nR_0},2^{nR_1},2^{nR_2},d_1,d_2)$ code with a probability of error $P_e^{(n)}$ that becomes arbitrarily small with the increase of the block length (see (8)). By Fano's inequality,

$$H(M_0,M_1,M_2|Y^n,S^n) \leq n(R_0+R_1+R_2)P_e^{(n)} + H(P_e^{(n)}) \triangleq n\varepsilon_n, \tag{16}$$

where clearly $\varepsilon_n \to 0$ as $P_e^{(n)} \to 0$. It hence follows that

$$H(M_1|Y^n,S^n) \leq H(M_0,M_1,M_2|Y^n,S^n) \leq n\varepsilon_n, \tag{17}$$

$$H(M_2|Y^n,S^n) \leq H(M_0,M_1,M_2|Y^n,S^n) \leq n\varepsilon_n, \tag{18}$$

$$H(M_1,M_2|Y^n,S^n) \leq H(M_0,M_1,M_2|Y^n,S^n) \leq n\varepsilon_n. \tag{19}$$

For the sake of brevity we focus here on the upper bound on R_1, while noting that all other upper bounds in (15) can be derived in a completely analogous manner, using the same auxiliary RV definition. It now follows that

$$\begin{aligned} nR_1 &= H(M_1) \\ &= H(M_1) + H(M_1|Y^n,S^n) - H(M_1|Y^n,S^n) \\ &\overset{(a)}{\leq} I(M_1;Y^n,S^n) + n\varepsilon_n \\ &\overset{(b)}{=} I(M_1;Y^n|S^n) + n\varepsilon_n \\ &\overset{(c)}{\leq} H(M_1|S^n,M_0,M_2) - H(M_1|Y^n,S^n,M_0,M_2) + n\varepsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^{n} I(M_1;Y_i|S^n,M_0,M_2,Y^{i-1}) + n\varepsilon_n \\ &\overset{(e)}{=} \sum_{i=1}^{n} [H(Y_i|S^n,X_2^n,M_0,M_2,Y^{i-1}) - H(Y_i|S^n,X_1^n,X_2^n,M_0,M_1,M_2,Y^{i-1})] + n\varepsilon_n \\ &\overset{(f)}{\leq} \sum_{i=1}^{n} [H(Y_i|X_{2,i},S_i,S_{i-d_1},S_{i-d_2},M_0,S^{i-d_1-1}) - H(Y_i|S^n,X_1^n,X_2^n,M_0,M_1,M_2,Y^{i-1})] + n\varepsilon_n \end{aligned}$$

$$\overset{(g)}{=} \sum_{i=1}^{n} [H(Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, M_0, S^{i-d_1-1}) - H(Y_i|X_{1,i}, X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, M_0, S^{i-d_1-1})] + n\varepsilon_n$$
$$\overset{(h)}{=} \sum_{i=1}^{n} I(X_{1,i}; Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, U_i) + n\varepsilon_n$$

where:

(a) follows from (17);

(b) follows from the fact that M_1 and S^n are independent;

(c) follows from the fact that M_1 and (M_0, M_2) are independent given S^n (first term), and the fact that conditioning reduces entropy (second term);

(d) follows from the mutual information chain rule;

(e) follows from the fact that that X_1^n is a deterministic function of (M_0, M_1, S^n) and X_2^n is a deterministic function of (M_0, M_2, S^n);

(f) follows from the fact that conditioning reduces entropy;

(g) follows from the fact that the channel output at time i depends only on the state S_i and the the inputs $X_{1,i}$ and $X_{2,i}$;

(h) follows by defining $U_i \triangleq (M_0, S^{i-d_1-1})$.

Note that the definition of the auxiliary RV U_i represents the common message and the common knowledge of the state sequence at time i (except for S_{i-d_1}), which, in fact, encompasses all common information shared by both encoders at this time instance. We can hence conclude that the rate R_1 must satisfy the following upper bound:

$$R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_{1,i}; Y_i|X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, U_i) + \varepsilon_n. \tag{20}$$

In a completely analogous manner it can be shown that

$$R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_{2,i}; Y_i|X_{1,i}, S_i, S_{i-d_2}, S_{i-d_1}, U_i) + \varepsilon_n, \tag{21}$$

$$R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i|S_i, S_{i-d_2}, S_{i-d_1}, U_i) + \varepsilon_n, \tag{22}$$

$$R_0 + R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i|S_i, S_{i-d_2}, S_{i-d_1}) + \varepsilon_n, \tag{23}$$

The upper bounds in (20)-(23) can also be rewritten by introducing a new time sharing RV Q, that is uniformly distributed over the set $\{1, 2, ..., n\}$. For example, the upper bound in (20) can be rewritten as

$$\begin{aligned} R_1 &\leq \frac{1}{n} \sum_{i=1}^{n} I(Y_Q; X_{1,Q}|X_{2,Q}, S_Q, S_{Q-d_2}, S_{Q-d_1}, U_Q, Q = i) + \varepsilon_n \\ &= I(Y_Q; X_{1,Q}|X_{2,Q}, S_Q, S_{Q-d_2}, S_{Q-d_1}, U_Q, Q) + \varepsilon_n. \end{aligned} \tag{24}$$

Repeating similar steps for all other upper bounds and denoting $X_1 \triangleq X_{1,Q}, X_2 \triangleq X_{2,Q}, Y \triangleq Y_Q, S \triangleq S_Q, \tilde{S}_1 \triangleq$

$S_{Q-d_1}, \tilde{S}_2 \triangleq S_{Q-d_2}$ and $U \triangleq (U_Q, Q)$, we get:

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2,U,S,\tilde{S}_1,\tilde{S}_2) + \varepsilon_n, \\ R_2 &\leq I(X_2;Y|X_1,U,S,\tilde{S}_1,\tilde{S}_2) + \varepsilon_n, \\ R_1 + R_2 &\leq I(X_1,X_2;Y|U,S,\tilde{S}_1,\tilde{S}_2) + \varepsilon_n, \\ R_0 + R_1 + R_2 &\leq I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2) + \varepsilon_n. \end{aligned}$$

Taking the limit as $n \to \infty$, one obtains the bounds as in (15).

To complete the proof of the converse it is left to show that the following Markov relations hold:

$$P(u|s,\tilde{s}_1,\tilde{s}_2) = P(u|\tilde{s}_1), \tag{25}$$

$$P(x_1|s,\tilde{s}_1,\tilde{s}_2,u) = P(x_1|\tilde{s}_1,u), \tag{26}$$

$$P(x_2|x_1,s,\tilde{s}_1,\tilde{s}_2,u) = P(x_2|\tilde{s}_1,\tilde{s}_2,u). \tag{27}$$

The proof of these relations is provided in Appendix A.

B. Achievability Proof

The proof of the achievability part relies on rate splitting, multiplexing coding and *joint* decoding. Fourier-Motzkin elimination is employed to reduce the number of inequalities induced by the error probability analysis.

To establish achievability, we need to show that for a fixed distribution $P(u|\tilde{s}_1)P(x_1|u,\tilde{s}_1)P(x_2|u,\tilde{s}_1,\tilde{s}_2)$ and rates (R_0,R_1,R_2) that satisfy the inequalities in (15), there exists a sequence of $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes such that $P_e^{(n)} \to 0$ as $n \to \infty$. Without loss of generality, we assume that the finite-state space is the set $\mathcal{S} = \{1,2,...,k\}$ and that the steady state probability satisfies $\pi(l) > 0$ for all $l \in \mathcal{S}$. Consider now the following encoding and decoding scheme.

1) Encoder 1: Encoder 1 constructs k codebooks $\mathcal{C}_0^{\tilde{s}_1}$ (the subscript 0 designates the message m_0) for all $\tilde{s}_1 \in \mathcal{S}$. Each codebook $\mathcal{C}_0^{\tilde{s}_1}$ comprises $2^{n_1(\tilde{s}_1)R_0(\tilde{s}_1)}$ codewords, where $n_1(\tilde{s}_1) = \big(P(\widetilde{S}_1 = \tilde{s}_1) - \epsilon'\big) \cdot n$ for some $\epsilon' > 0$. Each codeword in the codebook $\mathcal{C}_0^{\tilde{s}_1}$ is composed of $n_1(\tilde{s}_1)$ independent realizations of the random variable $U^{\tilde{s}_1}$, distributed identically according to $P(u^{\tilde{s}_1}|\widetilde{S}_1 = \tilde{s}_1)$.

Then, for every codeword $\mathcal{C}_0^{\tilde{s}_1}(i)$ in the codebook $\mathcal{C}_0^{\tilde{s}_1}$, $i \in \{1,2,....,2^{n_1(\tilde{s}_1)R_0(\tilde{s}_1)}\}$, Encoder 1 generates a codebook $\mathcal{C}_{i,1}^{\tilde{s}_1}$ comprises of $2^{n_1(\tilde{s}_1)R_1(\tilde{s}_1)}$ codewords. Each codeword in the codebook $\mathcal{C}_{i,1}^{\tilde{s}_1}$ is composed of $n_1(\tilde{s}_1)$ realizations of the random variable $X_1^{\tilde{s}_1}$ which are i.i.d according to $P(x_1^{\tilde{s}_1}|u^{\tilde{s}_1}, \widetilde{S}_1 = \tilde{s}_1)$. A message pair (M_0, M_1) is chosen according to a uniform distribution $\mathbb{P}(M_0 = m_0, M_1 = m_1) = 2^{-n(R_0+R_1)}$, where $m_0 \in \big\{1,2,...,2^{nR_0}\big\}$ and $m_1 \in \big\{1,2,...,2^{nR_1}\big\}$. Every message m_0 is split into k sub-messages $m_0 = \{m_{0,1}, m_{0,2}, ..., m_{0,k}\}$ (corresponding to each of the possible states), where the length of the sub-message $m_{0,\tilde{s}_1}$ is length $n_1(\tilde{s}_1)R_0(\tilde{s}_1)$. Hence, every message m_0 is specified by a k-dimensional vector. Similarly, every message m_1 is split into k sub-messages $m_1 = \{m_{1,1}, m_{1,2}, ..., m_{1,k}\}$ and is thus also specified by a k-dimensional vector. The length of the

sub-message $m_{1,\tilde{s}_1}$ is length $n_1(\tilde{s}_1)R_1(\tilde{s}_1)$.

For a fixed block length n, let $N_{\tilde{s}_1} \leq n$ be the number of times at which the feedback information at Encoder 1 regarding the channel state is $\widetilde{S}_1 = \tilde{s}_1$. Whenever the delayed CSI is $\widetilde{S}_1 = \tilde{s}_1$, Encoder 1 sends the next symbol from the codebook $\mathcal{C}_{i,1}^{\tilde{s}_1}$, where i is specified by the value of the sub-message of m_0 that corresponds to the state $\tilde{s}_1$. The construction of the codebooks for the sub-common and sub-private messages corresponding to some delayed CSI $\tilde{s}_1 = \ell$ is illustrated in Fig. 3. Encoder 1 repeats this construction for every $\ell \in \mathcal{S}$.

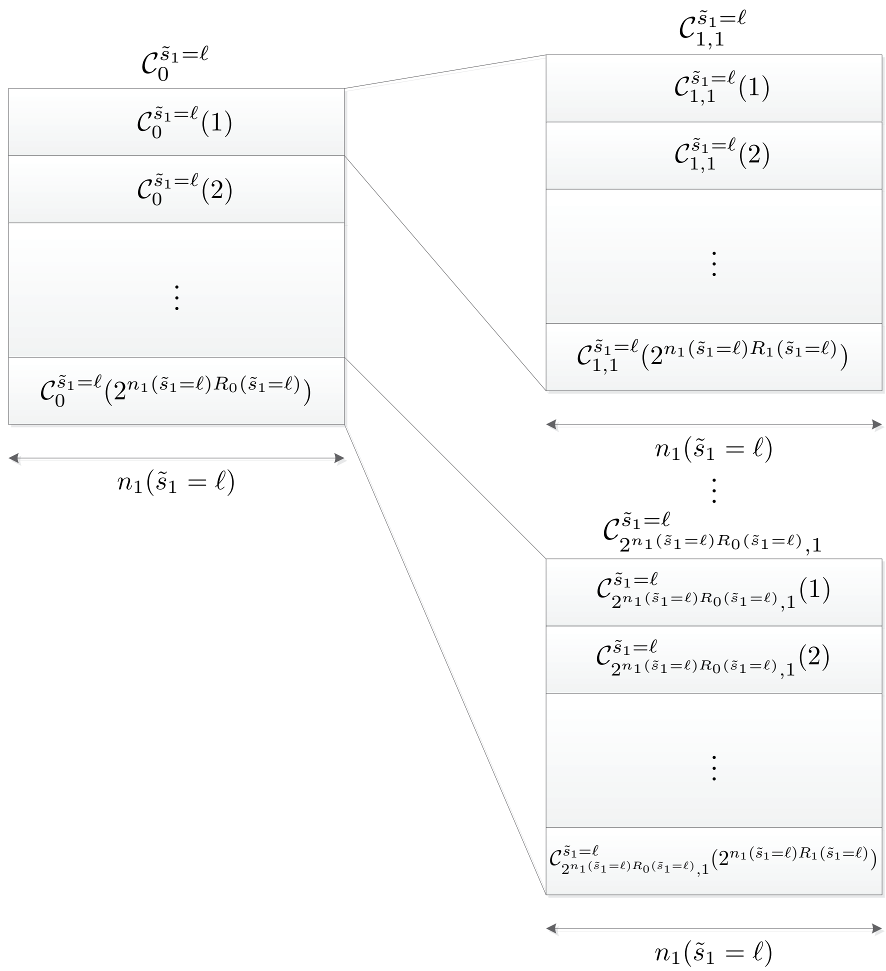


Fig. 3: Multiplexing coding with common message. Encoder 1's codebook is assembled from the common and private message codebooks. A common message codebook, $\mathcal{C}_0^{\tilde{s}_1=\ell}$, is generated for every sub-common message $m_{0,\ell}$, $\ell \in \mathcal{S}$, each containing $2^{n_1(\tilde{s}_1=\ell)R_0(\tilde{s}_1=\ell)}$ codewords. For every codeword in the ℓ-th common message codebook a private message codebook is generated, each containing $2^{n_1(\tilde{s}_1=\ell)R_1(\tilde{s}_1=\ell)}$ codewords.

2) Encoder 2: Encoder 2 first constructs k codebooks $\mathcal{C}_0^{\tilde{s}_1}$, for each $\tilde{s}_1 \in \mathcal{S}$, in an analogous manner to Encoder 1. Then, for every codeword $\mathcal{C}_0^{\tilde{s}_1}(i)$ in the codebook $\mathcal{C}_0^{\tilde{s}_1}$, $i \in \{1, 2, ..., 2^{n_1(\tilde{s}_1)R_0(\tilde{s}_1)}\}$, Encoder 2 generates k codebooks $\mathcal{C}_{i,2}^{\tilde{s}_1,\tilde{s}_2}$ for every $\tilde{s}_2 \in \mathcal{S}$. Each codebook is comprised of $2^{n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)}$ codewords, where $n_2(\tilde{s}_1, \tilde{s}_2) = \big(P(\widetilde{S}_1 = \tilde{s}_1, \widetilde{S}_2 = \tilde{s}_2) - \epsilon''\big) \cdot n$ for some $\epsilon'' > 0$. Each codeword in the codebook $\mathcal{C}_{i,2}^{\tilde{s}_1,\tilde{s}_2}$ is composed of $n_2(\tilde{s}_1, \tilde{s}_2)$ realizations

of the random variable $X_2^{\tilde{s}_1}$ which are i.i.d according to $P(x_2^{\tilde{s}_1,\tilde{s}_2}|u^{\tilde{s}_1},\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\tilde{s}_2)$. The mapping of the message m_0 is similar to the one defined for Encoder 1, while the mapping for m_2 is as follows. A message $m_2\in\{1,2,...,2^{nR_2}\}$ is chosen according to a uniform distribution $\mathbb{P}(M_2=m_2)=2^{-nR_2}$. Every message m_2 is split into $k\times k$ sub-messages $m_2=\{m_{2,1,1}\ ,\ m_{2,1,2}\ ,\ \ldots\ ,\ m_{2,k,k}\}$ (corresponding to each of the possible pairs of states). Each of these sub-messages is of length $n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)$. Hence, every message m_2 is specified by a k^2-dimensional vector.

For a fixed block length n, let $N_{\tilde{s}_1,\tilde{s}_2}$ be the number of times for which the feedback information at Encoder 2 regarding the channel state is $(\widetilde{S}_1,\widetilde{S}_2)=(\tilde{s}_1,\tilde{s}_2)$. Whenever the delayed CSI pair is $(\widetilde{S}_1,\widetilde{S}_2)=(\tilde{s}_1,\tilde{s}_2)$, Encoder 2 sends the next symbol from the codebook $\mathcal{C}_{i,2}^{\tilde{s}_1,\tilde{s}_2}$, where i is specified in a similar manner to its counterpart in Encoder 1.

3) Decoding process: The proposed decoding rule aims to achieve capacity by simultaneous decoding. Since full instantaneous CSI is assumed at the receiver, the information about $\widetilde{S}_1$ and $(\widetilde{S}_1,\widetilde{S}_2)$ used at the encoding stage is also available at the decoder.

Firs, we note that $N_{\tilde{s}_1}$ (respectively, $N_{\tilde{s}_1,\tilde{s}_2}$) is not necessarily equivalent to $n_1(\tilde{s}_1)$ (respectively, $n_2(\tilde{s}_1,\tilde{s}_2)$). Therefore, the decoder declares an error if $N_{\tilde{s}_1}<n_1(\tilde{s}_1)$ (respectively, $N_{\tilde{s}_1,\tilde{s}_2}<n_2(\tilde{s}_1,\tilde{s}_2)$), while the code is zero-filled if $N_{\tilde{s}_1}>n_1(\tilde{s}_1)$ (respectively, $N_{\tilde{s}_1,\tilde{s}_2}>n_2(\tilde{s}_1,\tilde{s}_2)$).

The decoding process is performed in blocks of size $n_1(\tilde{s}_1)$ corresponding to the delayed CSI $\widetilde{S}_1$. Upon receiving a block of channel outputs and states (Y^n,S^n), the decoder first demultiplexes it into outputs corresponding to the component codebooks of Encoder 1. The demultiplexing is done using the delayed CSI $\widetilde{S}_1$, which is known at the decoder. Then, the decoder simultaneously searches for $k+2$ unique sub-messages $\big(\hat{m}_{0\tilde{s}_1},\hat{m}_{1\tilde{s}_1},\hat{\mathbf{m}}_2(\tilde{s}_1)\big)$, that satisfy a certain typicality constraint to be specified next. Here $\hat{\mathbf{m}}_2(\tilde{s}_1)$ is the k-dimensional vector of sub-messages of m_2 for which $\widetilde{S}_1=\tilde{s}_1$ is fixed. The typicality constraint to be satisfied is:

$$\Big(U^{n_1(\tilde{s}_1)}(\hat{m}_{0\tilde{s}_1}),X_1^{n_1(\tilde{s}_1)}(\hat{m}_{0\tilde{s}_1},\hat{m}_{1\tilde{s}_1}),X_2^{n_1(\tilde{s}_1)}\big(\hat{m}_{0\tilde{s}_1},\hat{\mathbf{m}}_2(\tilde{s}_1)\big),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)},\widetilde{S}_2^{n_1(\tilde{s}_1)}\Big)\in\mathcal{T}_\epsilon^{(n_1(\tilde{s}_1))}(U,X_1,X_2,Y,S,\widetilde{S}_2), \tag{28}$$

for a given $\widetilde{S}_1$. By $\widetilde{S}_2^{n_1(\tilde{s}_1)}$ we refer to the sequence of $n_1(\tilde{s}_1)$ delayed channel states $\widetilde{S}_2$. Moreover, $U^{n_1(\tilde{s}_1)}(\hat{m}_{0\tilde{s}_1})$ represents the component codewords of the common sub-message $\hat{m}_{0\tilde{s}_1}$, while $X_1^{n_1(\tilde{s}_1)}(\hat{m}_{0\tilde{s}_1},\hat{m}_{1\tilde{s}_1})$ is the component codeword of the sub-message pair $(\hat{m}_{0\tilde{s}_1},\hat{m}_{1\tilde{s}_1})$. Finally, $X_2^{n_1(\tilde{s}_1)}(\hat{m}_{0\tilde{s}_1},\hat{\mathbf{m}}_2(\tilde{s}_1))$ is constructed by de-multiplexing the component codewords $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}$ using a common $\widetilde{S}_1=\tilde{s}_1$.

As mentioned earlier the decoding process is done in blocks of size $n_1(\tilde{s}_1)$, for each $\tilde{s}_1\in\mathcal{S}$. Thus, whenever the decoder succeeds in decoding each set of $k+2$ sub-messages (which consist of the k sub-messages in $\hat{\mathbf{m}}_2(\tilde{s}_1)$ as well as the sub-messages $\hat{m}_{0\tilde{s}_1}$ and $\hat{m}_{1\tilde{s}_1}$), it is immediately clear which triplet of messages $(\hat{m}_0,\hat{m}_1,\hat{m}_2)$ was sent, since each of the messages is uniquely determined by its sub-messages.

By error probability analysis we show that the probability of error, conditioned on a particular codeword being sent, goes to zero if the following conditions are met:

$$R_1'<I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{29}$$

$$\sum_{\ell\in\mathcal{S}_F} P(\ell)R_{2\ell} < \sum_{\ell\in\mathcal{S}_F} P(\ell)I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell), \tag{30}$$

$$R_1' + \sum_{\ell\in\mathcal{S}_F} P(\ell)R_{2\ell} < I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2) + \sum_{\ell\in\mathcal{S}_F} P(\ell)I(X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell), \tag{31}$$

$$R_0' + R_1' + R_2' < I(X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{32}$$

for every possible subset $\mathcal{S}_F \subseteq \mathcal{S}$. The subset $\mathcal{S}_F$ consists of all the states $\tilde{s}_2 \in \mathcal{S}$ for which a decoding error occurred and the reconstructed sub-message is incorrect (the subscript F stands for 'False'). Moreover, (29)-(32) are formulated using the following notations:

$$R_0(\tilde{s}_1) = R_0', \tag{33}$$

$$R_1(\tilde{s}_1) = R_1', \tag{34}$$

$$R_2(\tilde{s}_1, \tilde{s}_2=\ell) = R_{2\ell}, \tag{35}$$

$$\mathbb{P}(\widetilde{S}_2=\ell|\widetilde{S}_1=\tilde{s}_1) = P_\ell. \tag{36}$$

We also define

$$R_2(\tilde{s}_1) = \sum_{\ell=1}^{k} P_\ell R_{2\ell}, \tag{37}$$

and denote

$$R_2(\tilde{s}_1) = R_2'. \tag{38}$$

For the full detail and notation see Appendix B.

Next, using the Fourier-Motzkin elimination (FME), Appendix C shows that the set of inequalities of the form (30) and (31), which must hold for every $\mathcal{S}_F$, are equivalent to

$$R_2' < I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{39}$$

$$R_1' + R_2' < I(X_1,X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2). \tag{40}$$

Combining (29), (32), with (39) and (40), then in order for $P_e \to 0$ as $n \to \infty$, the following conditions must hold,

$$R_1(\tilde{s}_1) < I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{41}$$

$$R_2(\tilde{s}_1) < I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{42}$$

$$R_1(\tilde{s}_1) + R_2(\tilde{s}_1) < I(X_1,X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{43}$$

$$R_0(\tilde{s}_1) + R_1(\tilde{s}_1) + R_2(\tilde{s}_1) < I(X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2). \tag{44}$$

Considering the inequality in (41), the result can be extended to address all codebooks $C_{1,1}^{\tilde{s}_1}$, for every $\tilde{s}_1 \in \mathcal{S}$,

in the following way. We define

$$R_1 = \sum_{\tilde{s}_1 \in \mathcal{S}} \frac{n_1(\tilde{s}_1)}{n} R_1(\tilde{s}_1). \tag{45}$$

Using (45) we have

$$\begin{aligned} R_1 &\leq \sum_{\tilde{s}_1 \in \mathcal{S}} \frac{n_1(\tilde{s}_1)}{n} I(X_1; Y | X_2, U, S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2) \\ &= \sum_{\tilde{s}_1 \in \mathcal{S}} \big(P(\tilde{s}_1) - \epsilon'\big) I(X_1; Y | X_2, U, S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2) \\ &= I(X_1; Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) - \epsilon'', \end{aligned} \tag{46}$$

where $\epsilon'' = \epsilon' \cdot I(X_1; Y | X_2, U, S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2)$.

Performing a similar procedure for each of the bounds in (42)-(44), we get that in order to have $P_e \to 0$ as $n \to \infty$, the following rate constraints must be satisfied

$$R_1 < I(X_1; Y | X_2, U, S, \widetilde{S}_1, \widetilde{S}_2), \tag{47}$$

$$R_2 < I(X_2; Y | X_1, U, S, \widetilde{S}_1, \widetilde{S}_2), \tag{48}$$

$$R_1 + R_2 < I(X_1, X_2; Y | U, S, \widetilde{S}_1, \widetilde{S}_2), \tag{49}$$

$$R_0 + R_1 + R_2 < I(X_1, X_2; Y | S, \widetilde{S}_1, \widetilde{S}_2), \tag{50}$$

which holds by assumption.

Concluding, we have shown that if a rate triplet (R_0, R_1, R_2) is inside the rate region given in Theorem 1, then there exists a sequence of $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes such that $P_e^{(n)} \to 0$ as $n \to \infty$. This completes the proof of the achievability part.

IV. THE CAPACITY REGION OF THE FSM-MAC WITH PARTIALLY COOPERATIVE ENCODERS AND DELAYED TRANSMITTER CSI

In this section we state the capacity region of the FSM-MAC with partially cooperative encoders and delayed transmitter CSI, followed by it's proof. Without loss of generality, here we also assume that $d_1 \geq d_2$.

Theorem 2 *The capacity region of FSM-MAC with partially cooperative encoders, cooperation link capacities* C_{12} *and* C_{21}*, CSI at the decoder and asymmetrically delayed CSI at the encoders with delays* d_1 *and* d_2*, as shown in*

Fig. 1, is given by:

$$\mathcal{C}_{PC} = \bigcup_{P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)} \left\{ \begin{array}{rcl} R_1 & \leq & I(X_1;Y|X_2,U,S,\tilde{S}_1,\tilde{S}_2)+C_{12}, \\ R_2 & \leq & I(X_2;Y|X_1,U,S,\tilde{S}_1,\tilde{S}_2)+C_{21}, \\ R_1+R_2 & \leq & I(X_1,X_2;Y|U,S,\tilde{S}_1,\tilde{S}_2)+C_{12}+C_{21}, \\ R_1+R_2 & \leq & I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2). \end{array} \right\}, \tag{51}$$

where U is a RV with bounded cardinality and where the joint distribution of $(S,\widetilde{S}_1,\widetilde{S}_2)$ is specified in (5).

A. Converse

Given an achievable rate (R_1,R_2), we need to show that there exists a joint distribution of the form $P(s,\tilde{s}_1,\tilde{s}_2)P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)P(y|x_1,x_2,s)$ such that the inequalities (51) are satisfied. Since (R_1,R_2) is an achievable rate-pair, there exists an $(n,l,2^{nR_1},2^{nR_2},d_1,d_2)$ code with an arbitrarily small error probability $P_e^{(n)}$. By Fano's inequality,

$$H(M_1,M_2|Y^n,S^n) \leq n(R_1+R_2)P_e^{(n)} + H(P_e^{(n)}) \triangleq n\varepsilon_n, \tag{52}$$

(with some abuse of notation) where $\varepsilon_n \to 0$ as $P_e^{(n)} \to 0$. It hence follows that

$$H(M_1|Y^n,S^n) \leq H(M_1,M_2|Y^n,S^n) \leq n\varepsilon_n, \tag{53}$$

$$H(M_2|Y^n,S^n) \leq H(M_1,M_2|Y^n,S^n) \leq n\varepsilon_n. \tag{54}$$

As in the proof of Theorem 1, we focus on the upper bound on R_1 and note that the upper bounds on all other rates can be straightforwardly obtained in an analogous manner. For R_1 we have the following:

$$\begin{aligned}
nR_1 &= H(M_1) \\
&= H(M_1) + H(M_1|Y^n,S^n) - H(M_1|Y^n,S^n) \\
&\overset{(a)}{\leq} I(M_1;Y^n,S^n) + n\varepsilon_n \\
&\overset{(b)}{=} I(M_1;Y^n|S^n) + n\varepsilon_n \\
&\overset{(c)}{\leq} H(M_1|S^n,M_2) - H(M_1|V_1^\ell,V_2^\ell,Y^n,S^n,M_2) + n\varepsilon_n \\
&\overset{(d)}{=} I(M_1;V_1^\ell,V_2^\ell|S^n,M_2) + I(M_1;Y^n|V_1^\ell,V_2^\ell,S^n,M_2) + n\varepsilon_n \\
&\overset{(e)}{=} H(V_1^\ell|S^n,M_2) + H(V_2^\ell|V_1^\ell,S^n,M_2) + \sum_{i=1}^{n} I(M_1;Y_i|V_1^\ell,V_2^\ell,S^n,M_2,Y^{i-1}) + n\varepsilon_n \\
&\overset{(f)}{=} H(V_1^\ell|S^n,M_2) + \sum_{i=1}^{n} [H(Y_i|V_1^\ell,V_2^\ell,S^n,M_2,Y^{i-1}) - H(Y_i|V_1^\ell,V_2^\ell,S^n,M_1,M_2,Y^{i-1})] + n\varepsilon_n \\
&\overset{(g)}{\leq} H(V_1^\ell) + \sum_{i=1}^{n} [H(Y_i|V_1^\ell,V_2^\ell,S^n,X_2^n,M_2,Y^{i-1}) - H(Y_i|V_1^\ell,V_2^\ell,S^n,X_1^n,X_2^n,M_1,M_2,Y^{i-1})] + n\varepsilon_n
\end{aligned}$$

$$\stackrel{(h)}{\leq} \sum_{j=1}^{\ell} H(V_{1,j}|V_1^{j-1}) + \sum_{i=1}^{n}[H(Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, V_1^\ell, V_2^\ell, S^{i-d_1-1})$$

$$- H(Y_i|V_1^\ell, V_2^\ell, S^n, X_1^n, X_2^n, M_1, M_2, Y^{i-1})] + n\varepsilon_n$$

$$\stackrel{(i)}{\leq} \sum_{j=1}^{\ell} H(V_{1,j}) + \sum_{i=1}^{n}[H(Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, V_1^\ell, V_2^\ell, S^{i-d_1-1})$$

$$- H(Y_i|X_{1,i}, X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, V_1^\ell, V_2^\ell, S^{i-d_1-1})] + n\varepsilon_n$$

$$\stackrel{(j)}{\leq} \sum_{j=1}^{\ell} \log|V_{1,j}| + \sum_{i=1}^{n} I(X_{1,i}; Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, V_1^\ell, V_2^\ell, S^{i-d_1-1}) + n\varepsilon_n$$

$$\stackrel{(k)}{\leq} nC_{12} + \sum_{i=1}^{n} I(X_{1,i}; Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, U_i) + n\varepsilon_n$$

where:

(a) follows from (53);

(b) follows from the fact that M_1 and S^n are independent;

(c) follows from the fact that M_1 and M_2 are independent given S^n (first term), and the fact that conditioning reduces entropy (second term);

(d) follows from adding and subtracting the term $H(M_1|V_1^\ell, V_2^\ell, S^n, M_2)$;

(e) follows from the entropy and mutual information chain rules and the fact that V_1^ℓ and V_2^ℓ are fully determined given (S^n, M_1, M_2);

(f) follows from the fact that V_2^ℓ is a deterministic function of (M_2, V_1^ℓ);

(g) follows from the fact that conditioning reduces entropy (first term), the fact that X_1^n is a deterministic function of (M_1, V_1^ℓ, S^n) and that X_2^n is a deterministic function of (M_2, V_2^ℓ, S^n) (second and third term);

(h) follows form the entropy chain rule (first term) and the fact that conditioning reduces entropy (second term);

(i) follows from the fact that conditioning reduces entropy (first term) and the fact that that the channel output at time i depends only on the state S_i and the inputs $X_{1,i}$ and $X_{2,i}$ (third term);

(j) follows from the upper bound of entropy;

(k) follows from defining $U_i = (V_1^\ell, V_2^\ell, S^{i-d_1-1})$.

Note that we defined here the auxiliary RV U at time i as $U_i \triangleq (V_1^\ell, V_2^\ell, S^{i-d_1-1})$. Hence, it represents the information shared during the conference (i.e., the parts of the private messages available to both encoders) and the common knowledge of the states. As was the case for the common message setting (cf. Theorem 1 and Subsection III-A), U here represents all the common information available to both users.

Applying similar arguments to R_2 and $R_1 + R_2$ one can conclude that any achievable rate-pair (R_1, R_2) must satisfy the following inequalities:

$$R_1 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i}; Y_i|X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, U_i) + C_{12} + \varepsilon_n, \tag{55}$$

$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2,i};Y_i|X_{1,i},S_i,S_{i-d_2},S_{i-d_1},U_i) + C_{21} + \varepsilon_n, \tag{56}$$

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i},X_{2,i};Y_i|S_i,S_{i-d_2},S_{i-d_1},U_i) + C_{12} + C_{21} + \varepsilon_n, \tag{57}$$

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i},X_{2,i};Y_i|S_i,S_{i-d_2},S_{i-d_1}) + \varepsilon_n. \tag{58}$$

The expressions on the RHS of the inequalities in (55)-(58) represent empirical averages of mutual informations (taken over the code symbols). These inequalities can be alternatively represented by introducing a new time-sharing RV Q, uniformly distributed over $\{1,\ldots,n\}$, as in Subsection III-A. Starting again with the upper bound on R_1, this yields

$$\begin{aligned} R_1 &\leq \frac{1}{n}\sum_{i=1}^{n} I(Y_Q;X_{1,Q}|X_{2,Q},S_Q,S_{Q-d_2},S_{Q-d_1},U_Q,Q=i) + C_{12} + \varepsilon_n \\ &= I(Y_Q;X_{1,Q}|X_{2,Q},S_Q,S_{Q-d_2},S_{Q-d_1},U_Q,Q) + C_{12} + \varepsilon_n \end{aligned} \tag{59}$$

Applying the same procedure to the rest of the upper bounds, while denoting $X_1 \triangleq X_{1,Q}, X_2 \triangleq X_{2,Q}, Y \triangleq Y_Q, S \triangleq S_Q, \tilde{S}_1 \triangleq (S_{Q-d_1},Q), \tilde{S}_2 \triangleq S_{Q-d_2}$ and $U \triangleq (U_Q,Q)$, we get

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2,U,S,\tilde{S}_1,\tilde{S}_2) + C_{12} + \varepsilon_n, \\ R_2 &\leq I(X_2;Y|X_1,U,S,\tilde{S}_1,\tilde{S}_2) + C_{21} + \varepsilon_n, \\ R_1 + R_2 &\leq I(X_1,X_2;Y|U,S,\tilde{S}_1,\tilde{S}_2) + C_{12} + C_{21} + \varepsilon_n, \\ R_1 + R_2 &\leq I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2) + \varepsilon_n. \end{aligned}$$

To complete the proof of the converse it is left to show that the following Markov relations hold:

$$P(u|s,\tilde{s}_1,\tilde{s}_2) = P(u|\tilde{s}_1), \tag{60}$$

$$P(x_1|s,\tilde{s}_1,\tilde{s}_2,u) = P(x_1|\tilde{s}_1,u), \tag{61}$$

$$P(x_2|x_1,s,\tilde{s}_1,\tilde{s}_2,u) = P(x_2|\tilde{s}_1,\tilde{s}_2,u), \tag{62}$$

The proof of these relations is given in Appendix D.

It can thus be concluded, by taking the limit as $n \to \infty$, $P_e^{(n)} \to 0$, that the following holds

$$R_1 \leq I(X_1;Y|X_2,U,S,\tilde{S}_1,\tilde{S}_2) + C_{12}, \tag{63}$$

$$R_2 \leq I(X_2;Y|X_1,U,S,\tilde{S}_1,\tilde{S}_2) + C_{21}, \tag{64}$$

$$R_1 + R_2 \leq I(X_1,X_2;Y|U,S,\tilde{S}_1,\tilde{S}_2) + C_{12} + C_{21}, \tag{65}$$

$$R_1 + R_2 \leq I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2). \tag{66}$$

for some choice of joint distribution $P(s,\tilde{s}_1,\tilde{s}_2)P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)P(y|x_1,x_2,s)$.

B. Achievability

To prove the achievability of the capacity region, we need to show that for a fixed distribution of the form $P(u|\tilde{s}_1)P(x_1|u,\tilde{s}_1)P(x_2|u,\tilde{s}_1,\tilde{s}_2)$ and for (R_1,R_2) that satisfy the inequalities in (51), there exists a sequence of $(n,\ell,2^{nR_1},2^{nR_2},d_1,d_2)$ codes for which $P_e^{(n)}\to 0$ as $n\to\infty$.

The idea behind this proof is to convert the conferencing problem into a setting that corresponds to the FSM-MAC with common message considered in Section III, and then rely on its capacity result to show that the conferencing capacity region is, indeed, achievable. This is done by sharing as much as possible of the original private messages (m_1,m_2), through the conferencing links, in order to create a common message. The unshared parts of the original messages serve as the private messages in the new setting. By doing so, the coding scheme of the setting with a common message, as detailed in Section III, can be employed.

We start by defining:

$$\widetilde{R}_1=\min\{R_1,C_{12}\}, \tag{67}$$

$$\widetilde{R}_2=\min\{R_2,C_{21}\}. \tag{68}$$

With respect to these definitions, the inequalities in (51) can be rewritten as

$$\begin{aligned}
(R_1-\widetilde{R}_1) &\le I(X_1;Y|X_2,U,S,\tilde{S}_1,\tilde{S}_2),\\
(R_2-\widetilde{R}_2) &\le I(X_2;Y|X_1,U,S,\tilde{S}_1,\tilde{S}_2),\\
(R_1-\widetilde{R}_1)+(R_2-\widetilde{R}_2) &\le I(X_1,X_2;Y|U,S,\tilde{S}_1,\tilde{S}_2),\\
(\widetilde{R}_1+\widetilde{R}_2)+(R_1-\widetilde{R}_1)+(R_2-\widetilde{R}_2) &\le I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2).
\end{aligned} \tag{69}$$

In view of this representation, we construct a coding scheme by splitting the sets $\mathcal{M}_j=\{1,2,\ldots,2^{nR_j}\}$, for $j\in\{1,2\}$, into $2^{n\widetilde{R}_j}$ cells, each containing $2^{n(R_j-\widetilde{R}_j)}$ messages, and introducing the functions

$$c_1 \;:\; \mathcal{M}_1\to\{1,2,\ldots,2^{n\widetilde{R}_1}\}, \tag{70}$$

$$c_2 \;:\; \mathcal{M}_2\to\{1,2,\ldots,2^{n\widetilde{R}_2}\}, \tag{71}$$

$$e_1 \;:\; \mathcal{M}_1\to\{1,2,\ldots,2^{n(R_1-\widetilde{R}_1)}\}, \tag{72}$$

$$e_2 \;:\; \mathcal{M}_2\to\{1,2,\ldots,2^{n(R_2-\widetilde{R}_2)}\}. \tag{73}$$

That is, for every message m_j, where $j\in\{1,2\}$, c_j returns its cell number, $c_j(m_j)$, while e_j returns its index number, $e_j(m_j)$, inside the cell $c_j(m_j)$. For the sake of simplicity, we assume here that $2^{n\widetilde{R}_1}$, $2^{n\widetilde{R}_2}$, $2^{n(R_1-\widetilde{R}_1)}$ and $2^{n(R_2-\widetilde{R}_2)}$ are integers, although the same approach can be formalized for real numbers as well. Also note that the partitioning is deterministic.

Now, for every message pair (m_1, m_2) let us define the triplet (m_0', m_1', m_2') where

$$m_1' \triangleq e_1(m_1), \tag{74}$$

$$m_2' \triangleq e_2(m_2), \tag{75}$$

$$m_0' \triangleq \big(c_1(m_1), c_2(m_2)\big). \tag{76}$$

Note that the above definitions dictate that $m_1' \in \{1, 2, \ldots, 2^{n(R_1 - \widetilde{R}_1)}\}$, $m_2' \in \{1, 2, \ldots, 2^{n(R_2 - \widetilde{R}_2)}\}$ and $m_0' \in \{1, 2, \ldots, 2^{n\widetilde{R}_1}\} \times \{1, 2, \ldots, 2^{n\widetilde{R}_2}\}$. Since, by definition,

$$\widetilde{R}_1 \leq C_{12}, \tag{77}$$

$$\widetilde{R}_2 \leq C_{21}, \tag{78}$$

it is possible for Encoder 1 to transmit $c_1(m_1)$ to Encoder 2, and for Encoder 2 to transmit $c_2(m_2)$ to Encoder 1, via the respective conferencing links. Therefore, following the conferencing stage, both encoders know $\big(c_1(m_1), c_2(m_2)\big)$, which can be viewed as a common message, and henceforth denoted as m_0'; m_1' and m_2' are viewed as the new private messages.

The above setting can hence be viewed as a FSM-MAC with common message. The messages to be transmitted are given by the triplet (m_0', m_1', m_2'), where $m_0' \in \{1, 2, \ldots, 2^{n\widetilde{R}_1}\} \times \{1, 2, \ldots, 2^{n\widetilde{R}_2}\}$, $m_1' \in \{1, 2, \ldots, 2^{n(R_1 - \widetilde{R}_1)}\}$ and $m_1' \in \{1, 2, \ldots, 2^{n(R_1 - \widetilde{R}_1)}\}$, while (69) holds by assumption. By Theorem 1, it now immediately follows that the new message triplet (m_0', m_1', m_2') can be transmitted to the decoder with an arbitrary small probability of error. The decoder can, therefore, reliably reconstruct the message pair (m_1, m_2) and the rate-region (51) is hence achievable.

V. The Vector Gaussian FSM-MAC with Diagonal Channel Transfer Matrices, Conferencing and Delayed CSI

In this section we consider the vector Gaussian FSM-MAC with diagonal channel transfer matrices, partially cooperative encoders and delayed CSI. For every time instance $t \in \{1, \ldots, n\}$, the channel model in concern is given by:

$$\mathbf{Y}_t = \mathrm{G}_1(s_t)\mathbf{X}_{1,t} + \mathrm{G}_2(s_t)\mathbf{X}_{2,t} + \mathbf{Z}_t, \tag{79}$$

where $\big\{\mathrm{G}_1(s)\big\}_{s\in\mathcal{S}}$ and $\big\{\mathrm{G}_2(s)\big\}_{s\in\mathcal{S}}$ are $N \times N$ diagonal matrices, which are deterministic functions of the channel state $S = s$ (for simplicity, we henceforth omit the time index t). We denote the diagonal entries of these matrices by $g_{1,i}(s)$ and $g_{2,i}(s)$, respectively, for $i \in \{1, \ldots, N\}$. Moreover, we assume $\mathrm{G}_1(s), \mathrm{G}_2(s) \in \mathbb{C}^{N\times N}$ for every $s \in \mathcal{S}$. $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{C}^N$ and $\mathbf{Y} \in \mathbb{C}^N$ are the channel input vectors and the channel output vector, respectively. $\mathbf{Z}$ is a proper complex Gaussian vector, independent of $\mathbf{X}_1$ and $\mathbf{X}_2$ and distributed according to $\mathbf{Z} \sim \mathcal{CN}(0, \mathrm{I})$, where I is the identity matrix of dimensions $N \times N$. The input vector signals are assumed to satisfy the average power

constraints

$$\mathrm{tr}\big(\Sigma_{X_1X_1}\big) \le \mathcal{P}_1 \quad ; \quad \mathrm{tr}\big(\Sigma_{X_2X_2}\big) \le \mathcal{P}_2, \tag{80}$$

where we use the standard notation $\Sigma_{XY} = \mathbb{E}[\mathbf{X}\mathbf{Y}^\dagger]$, where $\mathrm{A}^\dagger$ denotes the conjugate transpose of the matrix A.

The motivation for examining the channel mode in (79) stems from the fact that it can be used to represent an OFDM-based communication system, employing single receive and transmit antennas. OFDM is an efficient technique for mitigating frequency selective fading, which are typical to modern wide-band communication systems (see, e.g., [1], [42]). The underlying idea behind OFDM is to split the channel's bandwidth into N separate sub-channels through which orthogonal signals are transmitted. By doing so, not only is the effect of intersymbol interference (ISI) dramatically reduced, but the transfer functions of each of the sub-channels boil down to multiplicative scalar gains. These gains are modeled by the diagonal entries of the channel matrices defined above. In this section we derive the maximization problem that defines the capacity region for the vector Gaussian channel in concern and convert it into a convex form. The solution of this convex maximization problem, which can be easily obtained using a numerical tool such as CVX [41], also yields the optimal power allocation strategy among the sub-channels, which is another essential factor in an OFDM-based transmission.

A. Capacity Region

Theorem 3 *The capacity region of the power-constrained vector Gaussian FSM-MAC with diagonal channel transfer matrices, partially cooperative encoders, cooperation link capacities C_{12} and C_{21}, delayed CSI and average power constraints $(\mathcal{P}_1, \mathcal{P}_2)$ is the union of all sets of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying*

$$R_1 \le \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s} K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^{N} \log\Big(1 + |g_{1,i}(s)|^2 \gamma_{1,i}(\tilde{s}_1)\Big) + C_{12}, \tag{81}$$

$$R_2 \le \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s} K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^{N} \log\Big(1 + |g_{2,i}(s)|^2 \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2)\Big) + C_{21}, \tag{82}$$

$$R_1 + R_2 \le \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s} K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^{N} \log\Big(1 + |g_{1,i}(s)|^2 \gamma_{1,i}(\tilde{s}_1) + |g_{2,i}(s)|^2 \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2)\Big) + C_{12} + C_{21}, \tag{83}$$

$$R_1 + R_2 \le \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s} K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^{N} \log\Big(1 + |g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) + |g_{2,i}(s)|^2 P_{2,i}(\tilde{s}_2, \tilde{s}_1) + 2 g_{1,i}(s) g_{2,i}^*(s) \sqrt{\big(P_{1,i}(\tilde{s}_1) - \gamma_{1,i}(\tilde{s}_1)\big)\big(P_{2,i}(\tilde{s}_1, \tilde{s}_2) - \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2)\big)}\Big), \tag{84}$$

where the union is taken over all $\{\gamma_{1,i}(\tilde{s}_1)\}_{i\in\{1,\ldots,N\},\tilde{s}_1\in\mathcal{S}}$, $\{\gamma_{2,i}(\tilde{s}_1,\tilde{s}_2)\}_{i\in\{1,\ldots,N\},(\tilde{s}_1,\tilde{s}_2)\in\mathcal{S}}$, $\{P_{1,i}(\tilde{s}_1)\}_{i\in\{1,\ldots,N\},\tilde{s}_1\in\mathcal{S}}$ *and* $\{P_{2,i}(\tilde{s}_1,\tilde{s}_2)\}_{i\in\{1,\ldots,N\},(\tilde{s}_1,\tilde{s}_2)\in\mathcal{S}}$ *that satisfy the constraints:*

$$\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{i=1}^{N}P_{1,i}(\tilde{s}_1)\leq\mathcal{P}_1, \tag{85}$$

$$\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{i=1}^{N}P_{2,i}(\tilde{s}_1,\tilde{s}_2)\leq\mathcal{P}_2, \tag{86}$$

$$0\leq\gamma_{1,i}(\tilde{s}_1)\leq P_{1,i}(\tilde{s}_1),\forall\ i\in\{1,\ldots,N\},\tilde{s}_1\in\mathcal{S}, \tag{87}$$

$$0\leq\gamma_{2,i}(\tilde{s}_1,\tilde{s}_2)\leq P_{2,i}(\tilde{s}_1,\tilde{s}_2),\forall\ i\in\{1,\ldots,N\},(\tilde{s}_1,\tilde{s}_2)\in\mathcal{S}^2. \tag{88}$$

The corresponding capacity region for the analogous setting with a common message can be obtained by taking:

$$\widetilde{R}_0=C_{12}+C_{21}\quad;\quad\widetilde{R}_1=\max\{0,R_1-C_{12}\}\quad;\quad\widetilde{R}_2=\max\{0,R_2-C_{21}\}, \tag{89}$$

where $\widetilde{R}_0$ denotes the common message rate, and $\widetilde{R}_1$ and $\widetilde{R}_2$ denote the rates of the private messages (according to the common message channel definition in Subsection II-A).

Corollary 4 *The capacity region of the power-constrained vector Gaussian FSM-MAC with diagonal channel transfer matrices, a common message, delayed CSI and average power constraints* $(\mathcal{P}_1,\mathcal{P}_2)$ *is the union of all sets of rate triplets* $(R_0,R_1,R_2)\in\mathbb{R}_+^3$ *satisfying*

$$R_1\leq\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2)\sum_{i=1}^{N}\log\Big(1+|g_{1,i}(s)|^2\gamma_{1,i}(\tilde{s}_1)\Big), \tag{90}$$

$$R_2\leq\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2)\sum_{i=1}^{N}\log\Big(1+|g_{2,i}(s)|^2\gamma_{2,i}(\tilde{s}_1,\tilde{s}_2)\Big), \tag{91}$$

$$R_1+R_2\leq\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2)\sum_{i=1}^{N}\log\Big(1+|g_{1,i}(s)|^2\gamma_{1,i}(\tilde{s}_1)+|g_{2,i}(s)|^2\gamma_{2,i}(\tilde{s}_1,\tilde{s}_2)\Big), \tag{92}$$

$$\begin{aligned}R_0+R_1+R_2\leq\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2)\sum_{i=1}^{N}\log\Big(&1+|g_{1,i}(s)|^2P_{1,i}(\tilde{s}_1)+|g_{2,i}(s)|^2P_{2,i}(\tilde{s}_2,\tilde{s}_1)\\&+2g_{1,i}(s)g_{2,i}^*(s)\sqrt{\big(P_{1,i}(\tilde{s}_1)-\gamma_{1,i}(\tilde{s}_1)\big)\big(P_{2,i}(\tilde{s}_1,\tilde{s}_2)-\gamma_{2,i}(\tilde{s}_1,\tilde{s}_2)\big)}\Big),\end{aligned} \tag{93}$$

where the union is taken over the same domain satisfying the constraints (85)-(88).

Note that the regions in Theorem 3 and Corollary 4 are both given in the form of a convex optimization problem, which can be solved efficiently using numerical tools. In the following proof we first derive a slightly different, yet equivalent, region for the Gaussian conferencing model; this original capacity region involves a non-convex optimization problem. Then, by defining new variables we convert the optimization problem into a convex one.

Proof: A straightforward extension of the result stated in Theorem 2 yields the capacity region of the *general* vector FSM-MAC with partially cooperative encoders and delayed CSI. The region is given by the closure of the set of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy (cf. (51))

$$\begin{aligned} R_1 &\leq I(\mathbf{X}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{U}, S, \tilde{S}_1, \tilde{S}_2) + C_{12}, \\ R_2 &\leq I(\mathbf{X}_2; \mathbf{Y}|\mathbf{X}_1, \mathbf{U}, S, \tilde{S}_1, \tilde{S}_2) + C_{21}, \\ R_1 + R_2 &\leq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|\mathbf{U}, S, \tilde{S}_1, \tilde{S}_2) + C_{12} + C_{21}, \\ R_1 + R_2 &\leq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|S, \tilde{S}_1, \tilde{S}_2), \end{aligned} \tag{94}$$

for some joint distribution of the form

$$P(\mathbf{u}|\tilde{s}_1)P(\mathbf{x}_1|\tilde{s}_1, \mathbf{u})P(\mathbf{x}_2|\tilde{s}_1, \tilde{s}_2, \mathbf{u}), \tag{95}$$

where $\mathbf{U}$ is an auxiliary random vector with bounded cardinality. Note that the structure of the conditional PDF in (95) implies the Markov relations:

$$\mathbf{U} - \widetilde{S}_1 - (S, \widetilde{S}_2), \tag{96}$$

$$\mathbf{X}_1 - (\mathbf{U}, \widetilde{S}_1) - (S, \widetilde{S}_2), \tag{97}$$

$$\mathbf{X}_2 - (\mathbf{U}, \widetilde{S}_1, \widetilde{S}_2) - (S, \mathbf{X}_1). \tag{98}$$

The proof of Theorem 3 consists of two main parts. First, we provide an outer bound for the general capacity region in (94). Then, by choosing a jointly Gaussian distribution for $(\mathbf{X}_1, \mathbf{U}, \mathbf{X}_2)$, we show that the upper bound is indeed achievable and thus characterizes the actual capacity region.

The outer bound for the capacity region is obtained by substituting the RVs $(\mathbf{X}_1, \mathbf{U}, \mathbf{X}_2)$ in (94) with appropriately chosen jointly Gaussian RVs $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$, which satisfy a certain Markovian relation. We conclude that the RVs $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$ indeed admit the desired Markov relation using the following lemma [40, Section 2, Theorem 1].

Lemma 5 *Let $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ be jointly Gaussian random vectors. Then $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ form a Markov chain $\mathbf{A} - \mathbf{B} - \mathbf{C}$ if and only if their covariance matrices satisfy:*

$$\Sigma_{AC} = \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BC}. \tag{99}$$

As before, we restrict the detailed derivation to the upper bound on R_1, while noting that all other bounds in (94) can be straightforwardly treated in an analogous manner. To this end, we rewrite the bound on R_1 as

$$R_1 \leq \sum_{\tilde{s}_1 \in \mathcal{S}} \pi(\tilde{s}_1) \sum_{\tilde{s}_2 \in \mathcal{S}} K^{d_1 - d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s \in \mathcal{S}} K^{d_2}(s, \tilde{s}_2) I(\mathbf{X}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{U}, s, \tilde{s}_1, \tilde{s}_2) \tag{100}$$

and proceed with upper bounding each of the mutual information terms in the sum. Consider:

$$
\begin{aligned}
I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2,\mathbf{U},s,\tilde{s}_1,\tilde{s}_2) &\overset{(a)}{=} h(\mathrm{G}_1(s)\mathbf{X}_1+\mathbf{Z}|\mathbf{U},s,\tilde{s}_1)-h(\mathbf{Z})\\
&\overset{(b)}{\leq} h(\mathrm{G}_1(s)\mathbf{X}_1+\mathbf{Z}|\mathbf{V},s,\tilde{s}_1)-h(\mathbf{Z})\\
&\overset{(c)}{\leq} h(\mathrm{G}_1(s)\mathbf{X}_1^G+\mathbf{Z}|\mathbf{V}^G,s,\tilde{s}_1)-h(\mathbf{Z}) \qquad (101)\\
&\leq \sum_{i=1}^{N}\Big\{h(g_{1,i}(s)X_{1,i}^G+Z_i,V_i^G|s,\tilde{s}_1)-h(V_i^G|\tilde{s}_1)\Big\}-h(\mathbf{Z})\\
&\overset{(d)}{=} \sum_{i=1}^{N}\Bigg\{\log\left((\pi e)^2\left[1+|g_{1,i}(s)|^2\left(\mathbb{E}[|X_{1,i}^G|^2|\tilde{s}_1]-\frac{\left|\mathbb{E}[X_{1,i}^G(V_i^G)^*|\tilde{s}_1]\right|^2}{\mathbb{E}[|V_i^G|^2|\tilde{s}_1]}\right)\right]\cdot\Big[\mathbb{E}[|V_i^G|^2|\tilde{s}_1]\Big]\right)\\
&\quad -\log\Big((\pi e)\mathbb{E}[|V_i^G|^2|\tilde{s}_1]\Big)-\log\big(\pi e\big)\Bigg\}\\
&\overset{(e)}{=} \sum_{i=1}^{N}\log\left(1+|g_{1,i}(s)|^2P_{1,i}(\tilde{s}_1)\left[1-\frac{\left|\mathbb{E}[X_{1,i}^G(V_i^G)^*|\tilde{s}_1]\right|^2}{P_{1,i}(\tilde{s}_1)\mathbb{E}[|V_i^G|^2|\tilde{s}_1]}\right]\right) \qquad (102)\\
&\overset{(f)}{=} \sum_{i=1}^{N}\log\Big(1+|g_{1,i}(s)|^2\beta_{1,i}(\tilde{s}_1)P_{1,i}(\tilde{s}_1)\Big) \qquad (103)
\end{aligned}
$$

where:

(a) follows from (79) and the Markov relations (96)-(98);

(b) follows from substituting the RV $\mathbf{U}$ for any given $\widetilde{S}_1=\tilde{s}_1$ with a new RV, $\mathbf{V}(\tilde{s}_1)\triangleq\mathbb{E}[\mathbf{X}_1|\mathbf{U},\tilde{s}_1]$. Note that this is in fact the optimal estimator in the minimum mean square error (MMSE) sense of $\mathbf{X}_1$ given $\mathbf{U}$, for each specified delayed CSI $\widetilde{S}_1=\tilde{s}_1$. Substituting $\mathbf{U}$ for any given $\widetilde{S}_1=\tilde{s}_1$ with $\mathbf{V}(\tilde{s}_1)$ increases the first entropy term in view of the fact that $\mathbf{V}(s_1)$ is a deterministic function of the pair $(\mathbf{U},\tilde{s}_1)$, whereas $h(\mathbf{Z})$ is not affected by the substitution. Moreover, one can easily confirm that $(\mathbf{X}_1,\mathbf{V},\mathbf{X}_2)$ satisfy the covariance condition (99), i.e., the relation

$$\Sigma_{X_1X_2}(\tilde{s}_1,\tilde{s}_2)=\Sigma_{X_1V}(\tilde{s}_1)\Sigma_{VV}^{-1}(\tilde{s}_1)\Sigma_{VX_2}(\tilde{s}_1,\tilde{s}_2) \qquad (104)$$

holds for every $(\tilde{s}_1,\tilde{s}_2)\in\mathcal{S}^2$. Note that the dependance of the covariance matrices on the states is induced by the Markov relation (96)-(98);

(c) follows from the maximum differential entropy lemma [43, Section 2.2] which states that the differential entropy, $h(\mathbf{X}|\mathbf{Y})$, for a pair of RVs $(\mathbf{X},\mathbf{Y})$ distributed according to $f_{\mathbf{XY}}(\mathbf{x},\mathbf{y})$, with covariance matrices Σ_{XX} and Σ_{YY}, is maximized for jointly Gaussian $(\mathbf{X},\mathbf{Y})$. Therefore, introducing the triplet $(\mathbf{X}_1^G,\mathbf{V}^G,\mathbf{X}_2^G)$ of zero-mean jointly Gaussian RVs with the same auto- and cross- covariance matrices as those of $(\mathbf{X}_1,\mathbf{V},\mathbf{X}_2)$, and replacing $(\mathbf{X}_1,\mathbf{V},\mathbf{X}_2)$ with $(\mathbf{X}_1^G,\mathbf{V}^G,\mathbf{X}_2^G)$, increases the first entropy term. Moreover, by Lemma 5, we conclude that the Gaussian triplet $(\mathbf{X}_1^G,\mathbf{V}^G,\mathbf{X}_2^G)$, for any given $(S,\widetilde{S}_1,\widetilde{S}_2)=(s,\tilde{s}_1,\tilde{s}_2)$, is Markov, i.e., $\mathbf{X}_1^G(\tilde{s}_1)-\mathbf{V}^G(\tilde{s}_1)-\mathbf{X}_2^G(\tilde{s}_1,\tilde{s}_2)$ holds.

(d) follows from explicitly evaluating each of the entropy terms;

(e) follows from defining $P_{1,i}(\tilde{s}_1) \triangleq \mathbb{E}[|X_{1,i}|^2|\tilde{s}_1]$ and $P_{2,i}(\tilde{s}_1,\tilde{s}_2) \triangleq \mathbb{E}[|X_{2,i}|^2|\tilde{s}_1,\tilde{s}_2]$ (note that these are in fact the i-th diagonal entries of the covariance matrices $\Sigma_{X_1^G X_1^G}(\tilde{s}_1)$ and $\Sigma_{X_2^G X_2^G}(\tilde{s}_1,\tilde{s}_2)$, respectively. For this reason, the constraints in (85)-(86) follow immediately from (80) by applying the law of total expectation);

(f) follows from defining

$$\bar{\beta}_{1,i}(\tilde{s}_1) = \left|\frac{\mathbb{E}[V_i^G(X_{1,i}^G)^*|\tilde{s}_1]}{\sqrt{\mathbb{E}[|X_{1,i}^G|^2|\tilde{s}_1]\mathbb{E}[|V_i^G|^2|\tilde{s}_1]}}\right|^2 = \frac{\left|E[V_i^G(X_{1,i}^G)^*|\tilde{s}_1]\right|^2}{P_{1,i}(\tilde{s}_1)\mathbb{E}[|V_i^G|^2|\tilde{s}_1]}, \tag{105}$$

$$\bar{\beta}_{2,i}(\tilde{s}_1,\tilde{s}_2)=\left|\frac{\mathbb{E}[V_i^G(X_{2,i}^G)^*|\tilde{s}_1,\tilde{s}_2]}{\sqrt{\mathbb{E}[|X_{2,i}^G|^2|\tilde{s}_1,\tilde{s}_2]\mathbb{E}[|V_i^G|^2|\tilde{s}_1]}}\right|^2=\frac{\left|E[V_i^G(X_{2,i}^G)^*|\tilde{s}_1,\tilde{s}_2]\right|^2}{P_{2,i}(\tilde{s}_1,\tilde{s}_2)\mathbb{E}[|V_i^G|^2|\tilde{s}_1]}, \tag{106}$$

where we use the notation $\bar{\alpha} = 1-\alpha$, $\alpha \in \mathbb{R}$.

Note that since $\bar{\beta}_{1,i}(\tilde{s}_1)$ (respectively, $\bar{\beta}_{2,i}(\tilde{s}_1,\tilde{s}_2)$) is defined to be the squared correlation coefficient between $X_{1,i}^G$ (respectively, $X_{2,i}^G$) and V_i^G for a given delayed CSI $\widetilde{S}_1 = \tilde{s}_1$ (respectively, delayed CSI pair $(\widetilde{S}_1,\widetilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)$), we have that $\beta_{1,i}(\tilde{s}_1), \beta_{2,i}(\tilde{s}_1,\tilde{s}_2) \in [0,1]$ for every $i \in \{1,\dots,N\}$. The upper bounds on R_2, and both upper bounds on the sum rate R_1+R_2 are constructed in a similar manner.

Next, we show that the upper bounds are also achievable. In order to do so, we take $(\mathbf{X}_1,\mathbf{U},\mathbf{X}_2)$ to be zero-mean jointly Gaussian RVs that admit the Markov relations (96)-(98), and for which the auto- and cross- covariance matrices $\Sigma_{X_1X_1}(\tilde{s}_1)$, $\Sigma_{X_2X_2}(\tilde{s}_1,\tilde{s}_2)$, $\Sigma_{UU}(\tilde{s}_1)$, $\Sigma_{X_1U}(\tilde{s}_1)$ and $\Sigma_{X_2U}(\tilde{s}_1,\tilde{s}_2)$ are diagonal for every $(\tilde{s}_1,\tilde{s}_2) \in \mathcal{S}^2$. Specifically, we take

$$\Sigma_{X_1X_1}(\tilde{s}_1) = \mathrm{diag}\Big(\big\{P_{1,i}(\tilde{s}_1)\big\}_{i=1}^N\Big), \tag{107}$$

$$\Sigma_{X_2X_2}(\tilde{s}_1,\tilde{s}_2) = \mathrm{diag}\Big(\big\{P_{2,i}(\tilde{s}_1,\tilde{s}_2)\big\}_{i=1}^N\Big), \tag{108}$$

and denote the diagonal entries of the three remaining covariance matrices by $\sigma_{U_i}^2(\tilde{s}_1)$, $\mathbb{E}[X_{1,i}U_i^*|\tilde{s}_1]$ and $\mathbb{E}[X_{2,i}U_i^*|\tilde{s}_1,\tilde{s}_2]$, respectively. Moreover, $(\mathbf{X}_1,\mathbf{U},\mathbf{X}_2)$ are chosen to have the same entry-wise correlations as $(\mathbf{X}_1^G,\mathbf{V}^G,\mathbf{X}_2^G)$, that is

$$\frac{\left|E[U_iX_{1,i}^*|\tilde{s}_1]\right|^2}{P_{1,i}(\tilde{s}_1)\sigma_{U_i}^2(\tilde{s}_1)} = \bar{\beta}_{1,i}(\tilde{s}_1), \tag{109}$$

$$\frac{\left|E[U_iX_{2,i}^*|\tilde{s}_1,\tilde{s}_2]\right|^2}{P_{2,i}(\tilde{s}_1,\tilde{s}_2)\sigma_{U_i}^2(\tilde{s}_1)} = \bar{\beta}_{2,i}(\tilde{s}_1,\tilde{s}_2). \tag{110}$$

The upper bounds are achieved by this choice of distribution and notations. We present the calculation only for R_1. As in (101), using the channel model and the Markov relations, we have that:

$$\begin{aligned} I(\mathbf{X}_1;\mathbf{Y}|\mathbf{X}_2,\mathbf{U},s,\tilde{s}_1,\tilde{s}_2) &= h(\mathrm{G}_1(s)\mathbf{X}_1+\mathbf{Z}|\mathbf{U},s,\tilde{s}_1) - h(\mathbf{Z}) \\ &= h(\mathrm{G}_1(s)\mathbf{X}_1+\mathbf{Z},\mathbf{U}|s,\tilde{s}_1) - h(\mathbf{U}|\tilde{s}_1) - h(\mathbf{Z}). \end{aligned} \tag{111}$$

Clearly

$$h(\mathbf{Z}) = \log\left((\pi e)^N\right), \tag{112}$$

$$h(\mathbf{U}|\tilde{s}_1) = \log\left((\pi e)^N \prod_{i=1}^{N} \sigma^2_{U_i}(\tilde{s}_1)\right) \tag{113}$$

Therefore, it is left to evaluate

$$h(\mathrm{G}_1(s)\mathbf{X}_1 + \mathbf{Z}, \mathbf{U}|s, \tilde{s}_1) = \log\left((\pi e)^{2N}\left|\widetilde{\Sigma}(s, \tilde{s}_1)\right|\right), \tag{114}$$

where $\widetilde{\Sigma}(s, \tilde{s}_1)$ is a block matrix of the structure

$$\widetilde{\Sigma}(s, \tilde{s}_1) = \begin{pmatrix} \mathrm{G}_1(s)\Sigma_{X_1X_1}(\tilde{s}_1)\mathrm{G}_1^{\dagger}(s) + \mathrm{I} & \mathrm{G}_1(s)\Sigma_{X_1U}(\tilde{s}_1) \\ \Sigma^{\dagger}_{X_1U}(\tilde{s}_1)\mathrm{G}_1^{\dagger}(s) & \Sigma_{UU}(\tilde{s}_1) \end{pmatrix}. \tag{115}$$

After some algebra it can be shown that:

$$\begin{aligned} \left|\widetilde{\Sigma}(s, \tilde{s}_1)\right| &= \left(\prod_{i=1}^{N}\left[|g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) + 1\right]\right) \cdot \left(\prod_{i=1}^{N}\left[\sigma^2_{U_i}(\tilde{s}_1) - \frac{|g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1)\bar{\beta}_{1,i}(\tilde{s}_1)\sigma^2_{U_i}(\tilde{s}_1)}{|g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) + 1}\right]\right) \\ &= \left(\prod_{i=1}^{N}\sigma^2_{U_i}(\tilde{s}_1)\right) \cdot \left(\prod_{i=1}^{N}\left[|g_{1,i}(s)|^2 \beta_{1,i}(\tilde{s}_1) P_{1,i}(\tilde{s}_1) + 1\right]\right). \end{aligned} \tag{116}$$

Substituting (116) along with (112) and (113) into (111) and summing the mutual information terms over all state triplets $(S, \widetilde{S}_1, \widetilde{S}_2) = (s, \tilde{s}_1, \tilde{s}_2)$, we achieve the upper bound for R_1. In a similar manner, all other upper bounds can be shown to be achievable. This characterizes the maximization problem defining the capacity region for the diagonal vector Gaussian FSM-MAC with partially cooperative encoders and delayed CSI. Note that through this proof we have shown the optimality of the Gaussian multivariate input distribution for this model.

The cautious reader must have noticed, however, that the obtained maximization problem is not convex. In order to convert it to a convex maximization problem, we substitute

$$\gamma_{1,i}(\tilde{s}_1) = \beta_{1,i}(\tilde{s}_1)P_{1,i}(\tilde{s}_1), \ \forall \tilde{s}_1 \in \mathcal{S} \tag{117}$$

$$\gamma_{2,i}(\tilde{s}_1, \tilde{s}_2) = \beta_{2,i}(\tilde{s}_1, \tilde{s}_2)P_{2,i}(\tilde{s}_1, \tilde{s}_2), \ \forall (\tilde{s}_1, \tilde{s}_2) \in \mathcal{S}^2. \tag{118}$$

for every $i \in \{1, \ldots, N\}$. This substitution yields the rate bounds given in (81)-(84) and concludes the proof. ■

B. Two-State Scalar AWGN Channel Example

To gain some intuition on the capacity region of the MAC with partially cooperative encoders and delayed CSI we now consider the scalar Gaussian channel with only two possible states. The scalar channel corresponds to taking $N = 1$ in the diagonal vector channel definition described in (79). We denote the two possible channel states by G and B (where G stands for ‘Good’ and B for ‘Bad’), thus $\mathcal{S} = \{G, B\}$. The two states differ in their associated channel gains. When $S = G$, the gains are $g_1(s = G) = g_2(s = G) \triangleq g_G$, whereas when $S = B$ the

gains are $g_1(s = B) = g_2(s = B) \triangleq g_B$. We assume without loss of generality that $g_G > g_B$. The Markov model of the state process is illustrated in Fig. 4.

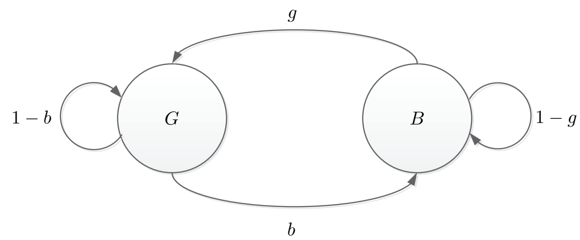


Fig. 4: Two-state AWGN channel.

The state process is specified by the the transition probability matrix:

$$K = \begin{pmatrix} P(G|G) & P(B|G) \\ P(G|B) & P(B|B) \end{pmatrix} = \begin{pmatrix} 1-b & b \\ g & 1-g \end{pmatrix}, \tag{119}$$

which induces the following stationary distribution:

$$\pi = \begin{pmatrix} \pi(G) & \pi(B) \end{pmatrix} = \begin{pmatrix} \frac{g}{g+b} & \frac{b}{g+b} \end{pmatrix}. \tag{120}$$

We start by examining the impact of the cooperation links capacities, C_{12} and C_{21}, on the capacity regions. We particularize here to the case of symmetric CSI delays, i.e. $d_1 = d_2 \triangleq d$. Note that since $d_1 = d_2$, it immediately follows that $\widetilde{S}_1 = \widetilde{S}_2 \triangleq \widetilde{S}$. The capacity region is presented in Fig. 5 for three different cases: (a) the case of symmetrical capacities, i.e., $C_{12} = C_{21}$, (b) the case of a single cooperation link, i.e., $C_{12} \geq C_{21} = 0$ and (c) the case of one infinite cooperation link, i.e., $C_{12} < C_{21} = \infty$. This is done by numerically solving the maximization problem from Theorem 3 for the above three cases using CVX [41]. Throughout this example we assume $\mathcal{P}_1 = \mathcal{P}_2 = 10$, $g_B = 0.01$, $g_G = 1$, $g = b = 0.1$ and $d = 2$ (results of similar nature were also observed for $g_B = 0.2$ and $g_B = 0.3$).

Note that in Fig. 5(a), which presents the region for the symmetrical case, as $C_{12} = C_{21}$ grows without bound, the capacity region increases and eventually takes the shape of a triangle. This is since the first three constraints on the rates (R_1, R_2), as given by (81)-(83), grow without bound as well, and thus the binding constraint is the sum-rate constraint of (84).For the single cooperation link case in Fig. 5(b), the value of R_2 remains fixed as C_{12} grows. This is due to the fact that the constraint on R_2 in (82) does not change with C_{12} and stays fixed at approximately 0.9642. Finally, for Fig. 5(c), which presents the case of infinite cooperation link capacity $C_{21} = \infty$, we have that the constraint on R_2 in (82) and the first constraint on the sum rate in (83), are both redundant. Hence, the only meaningful constraint on R_2 is (84), which does not involve C_{21} (or C_{12}).

Next, we demonstrate the fact that the capacity region of this setting grows as the cooperation link capacities grow, regardless of the specific assumptions on the relation between the delays of the CSI available at the encoders. In order to do so, we present the maximum sum rate versus the cooperation link capacities for three different

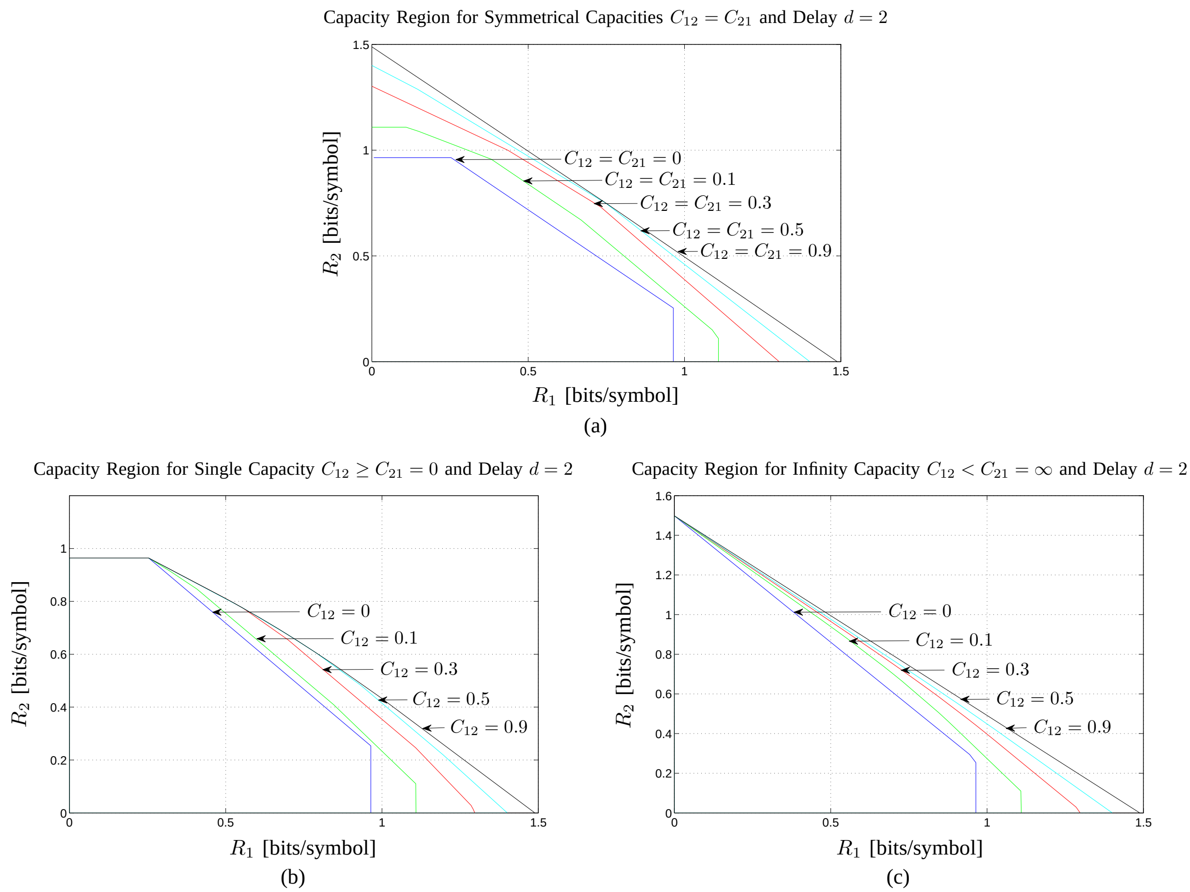
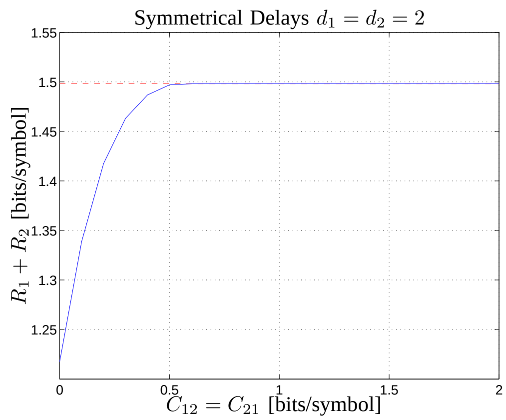


Fig. 5: Capacity region for the two-states AWGN-MAC: (a) symmetrical, $C_{12} = C_{21}$; (b) single cooperation link, $C_{12} \geq C_{21} = 0$; (c) infinite cooperation link, $C_{12} < C_{21} = \infty$.

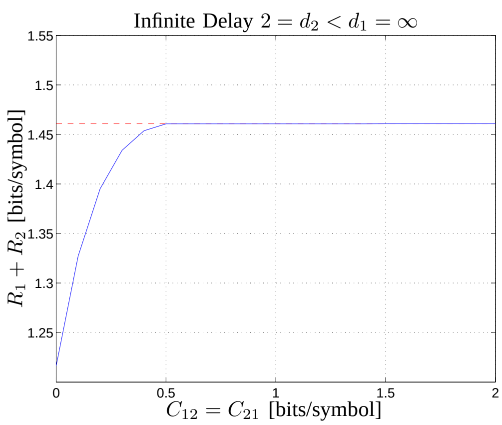
possible relations between the delays: (a) $d_1 = d_2 = 2$, (b) $2 = d_2 < d_1 = \infty$ and (c) $2 = d_1 \geq d_2 = 0$. For all three cases we assume $C_{12} = C_{21}$ and use the same values of the channel parameters as before. The curves are shown in Figs. 6(a)-(c).

As expected, the sum rate for the case of asymmetrical delays (which has the best CSI properties of the three) reaches the highest value as the capacities grow, whereas the sum rate for the infinite delay case (which has the worst CSI properties) reaches the lowest value. Moreover, we note the correspondence between Fig. 6(a) and Fig. 5(a) (both corresponding to the case of symmetrical delays and equal cooperation link capacities). This manifests itself in the fact that in both figures, as $C_{12} = C_{21}$ grow without bound, the sum rate approaches its maximal value, which is approximately 1.5 bits per symbol.

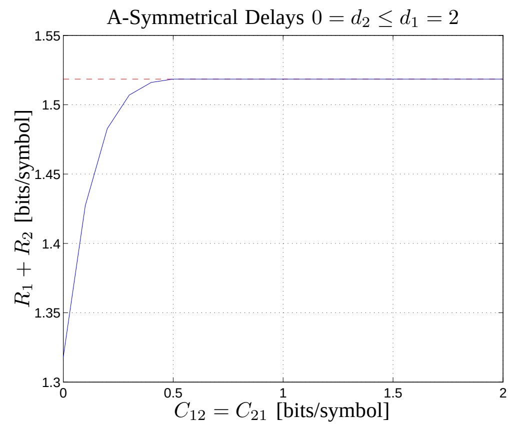
Another interesting aspect of the Gaussian channel example is the effect of signal to noise ration (SNR) on the correlations between the auxiliary RV, U, and the RVs X_1 and X_2. These correlations are associated with the degree of cooperation used in the scheme. We assume $P_1 = P_2 \triangleq P$ and $g_1 = g_2 = 1$, thus the SNR in fact equals to



(a)



(b)



(c)

Fig. 6: The sum rate versus the cooperation link capacities $C_{12} = C_{21}$ for three different cases of delayed CSI: (a) symmetrical delays, $d_2 = d_2 = 2$; (b) infinite delay, $2 = d_2 < d_1 = \infty$; (c) asymmetrical delays, $0 = d_2 \leq d_1 = 2$. The dashed line corresponds to the case where $C_{12} = C_{21} = \infty$.

the transmission power P. In order to examine the effect SNR on the correlations we restrict ourselves to the case where $|\mathcal{S}| = 1$, i.e., a single and constant channel state [37]. Throughout this analysis we use the same notations and expressions for the rate bounds as in [37]. Note that for the case where $|\mathcal{S}| = 1$ the original maximization problem turns out to be concave; thus, no transformation is needed. The only variables in the original maximization problem are β_1 and β_2, which are defined through

$$\sqrt{1-\beta_1} = \left| \frac{\mathbb{E}[UX_1]}{\sqrt{\mathbb{E}[X_1^2]\mathbb{E}[U^2]}} \right| \triangleq \rho_1, \ \ \sqrt{1-\beta_2} = \left| \frac{\mathbb{E}[UX_2]}{\sqrt{\mathbb{E}[X_2^2]\mathbb{E}[U^2]}} \right| \triangleq \rho_2. \tag{121}$$

We consider the case of symmetrical cooperation link capacities, i.e., $C_{12} = C_{21}$. By the symmetry of the maximization problem in (β_1, β_2), optimality is achieved when $\beta_1 = \beta_2$. For this reason we use the notation $\beta_1 = \beta_2 \triangleq \beta$ and plot a single curve representing both correlations (which are calculated directly from β according to (121)). The numerical results are shown in Fig. 7. The blue and green dashed lines designate the asymptotic

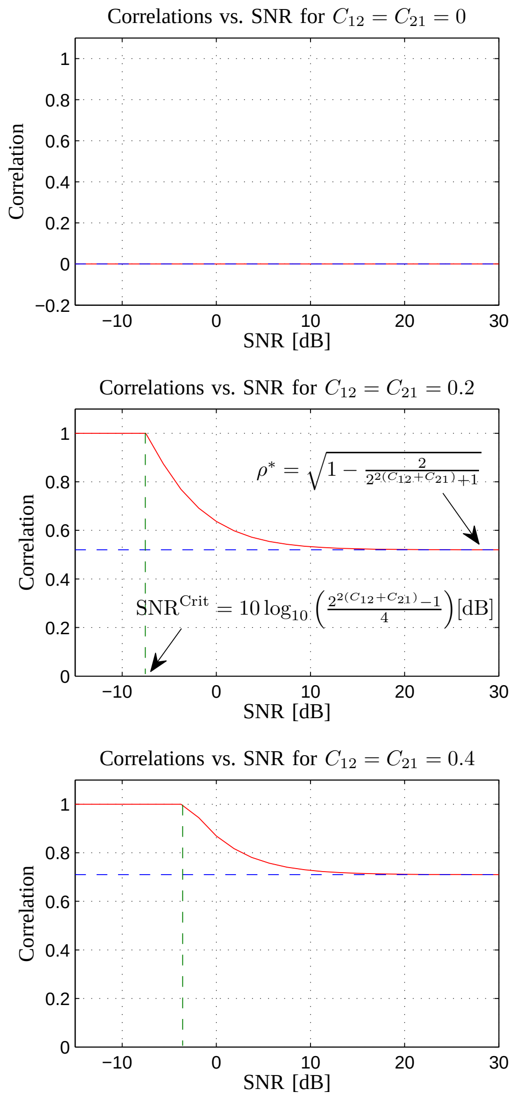


Fig. 7: Correlation as a function of SNR for different values of the capacities $C_{12} = C_{21}$.

value of the correlation and the critical SNR in which the correlation drops form unity, respectively. Results are shown for six different values of $C_{12} = C_{21}$.

Although the effect of the SNR on the correlations could not be calculated analytically, we use asymptotic

evaluations in order to gain some additional insight. Namely, we demonstrate that the optimal correlation admits

$$\rho^* = \begin{cases} 1 & , \quad \mathrm{SNR} \leq \mathrm{SNR}^{\mathrm{Crit}} \\ \sqrt{1 - \frac{2}{2^{2(C_{12}+C_{21})}+1}} & , \quad \mathrm{SNR} \to \infty \end{cases}, \tag{122}$$

where $\mathrm{SNR}^{\mathrm{Crit}} = 10\log_{10}\left(\frac{2^{2(C_{12}+C_{21})}-1}{4}\right)$[dB].

We start by justifying the observation that the correlation approaches 1 for small SNR values. For some positive value of $C_{12} = C_{21}$, and for and for $P_1, P_2 \ll 1$, consider (cf. (83)-(84)):

$$\begin{aligned} R_1 + R_2 \leq \min\Big\{ & \tfrac{1}{2}\log\Big(1+\beta(g_1^2P_1+g_2^2P_2)\Big)+C_{12}+C_{21} \quad , \quad \tfrac{1}{2}\log\Big(1+g_1^2P_1+g_2^2P_2+2g_1g_2\sqrt{\bar{\beta}^2P_1P_2}\Big) \Big\} \\ =& \frac{1}{2}\log\Big(1+g_1^2P_1+g_2^2P_2+2g_1g_2\sqrt{\bar{\beta}^2P_1P_2}\Big) \end{aligned} \tag{123}$$

Now note that that the last term in (123) is maximized for $\beta^* = 0$, which, in turn, implies that the correlation is equal to unity. As shown in Fig. 7, for smaller values of SNR the correlation is indeed higher, as if the scheme aims to compensate for the low SNR using cooperation.

The asymptotic evaluation for low SNRs is valid up to some critical SNR value in which the correlation drops from its maximal value of 1. We calculate this critical SNR next. We denote the SNR value of interest by $\mathrm{SNR}^{\mathrm{Crit}}$ and define it as

$$\mathrm{SNR}^{\mathrm{Crit}} = \sup\ \{P : \rho^*(P) = 1\}. \tag{124}$$

In order to calculate $\mathrm{SNR}^{\mathrm{Crit}}$ we restrict the analysis to the segment of SNRs in which the correlation is maximal (or equivalently, $\beta^* = 0$) and consider (123) taken for $P_1 = P_2 = P$ and $g_1 = g_2 = 1$ in. As shown in (123), when $\beta^* = 0$ and $P = 0$, the second logarithm achieves the minimum between the two terms. Fixing $\beta^* = 0$ and increasing P increases the second logarithm in (123) while the first term remains unchanged and equals to $C_{12} + C_{21}$. As long as

$$\frac{1}{2}\log\Big(1+2P+2\bar{\beta}P\Big)\Big|_{\beta=0} < C_{12}+C_{21} \tag{125}$$

holds, the optimum in achieved for $\beta^* = 0$. However, when (125) is no longer valid, the optimal value of β must vary from 0. Thus, calculating $\mathrm{SNR}^{\mathrm{Crit}}$ reduces to solving the following equation:

$$\frac{1}{2}\log\Big(1+2P+2\bar{\beta}P\Big)\Big|_{\beta=0} = C_{12}+C_{21}, \tag{126}$$

yielding,

$$\mathrm{SNR}^{\mathrm{Crit}} = \frac{2^{2(C_{12}+C_{21})}-1}{4}. \tag{127}$$

The value of $\mathrm{SNR}^{\mathrm{Crit}}$ [dB] is represented by the perpendicular dashed green line in the plots shown in Fig. 7; again the numerical calculations meet the analytical results. Note that as the capacities $C_{12} = C_{21}$ grow, so does the value of $\mathrm{SNR}^{\mathrm{Crit}}$, and hence the transition between the low- and high-SNR regimes occurs at a later stage.

As the SNR grows, the correlation asymptotically approaches some value in the interval $(0, 1)$. In order to find

this asymptotic correlation, we present the following analysis for the high-SNR regime (assuming $P_1, P_2 \gg 1$). We start by excluding $\beta^* = 0$ as a possible solution for this case (a fact which will be used subsequently). Fixing $C_{12} = C_{21}$ and substituting $\beta = 0$ into the sum-rate bounds on $R_1 + R_2$ yields (cf. (83)-(84)):

$$\begin{aligned} R_1 + R_2 &\leq \min\left\{ \; C_{12} + C_{21} \quad , \quad \tfrac{1}{2}\log\left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{P_1 P_2}\right) \; \right\} \\ &\overset{(a)}{=} C_{12} + C_{21} \end{aligned} \tag{128}$$

where (a) follows form the fact that $P_1, P_2 \gg 1$.
Thus we get that for an infinite SNR, by taking $\beta = 0$ the sum rate is bounded by the sum of capacities. However, since $C_{12} + C_{21}$ is a constant which does not depend on the powers P_1 and P_2, we conclude that β^* cannot be equal to zero.

Finally, assuming $\beta^* > 0$ we calculate β^* by using some approximations which are easily justified at high SNR. First, note that the first logarithm in (123) is monotonically increasing in β whereas the second is monotonically decreasing in β. This implies that the optimum is achieved at the value of β in which the functions intersect, that is

$$\frac{1}{2}\log\left(1 + \beta(g_1^2 P_1 + g_2^2 P_2)\right) + C_{12} + C_{21} = \frac{1}{2}\log\left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2}\right). \tag{129}$$

Using the fact that for high SNR we have:

$$\begin{aligned} \frac{1}{2}\log\left(1 + \beta(g_1^2 P_1 + g_2^2 P_2)\right) + C_{12} + C_{21} &\approx \frac{1}{2}\log\left(\beta(g_1^2 P_1 + g_2^2 P_2)\right) + C_{12} + C_{21}, \\ \frac{1}{2}\log\left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2}\right) &\approx \frac{1}{2}\log\left(g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2}\right), \end{aligned}$$

the equation in (129) reduces to:

$$\beta(g_1^2 P_1 + g_1^2 P_2) 2^{2(C_{12}+C_{21})} = g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2} \; .$$

In order to further simplify the analysis we again assume a unit channel gain, that is $g_1 = g_2 = 1$. After some algebra we obtain that the intersection point is given by

$$\beta^* = \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{2^{2(C_{12}+C_{21})}(P_1 + P_2) + 2\sqrt{P_1 P_2}} \; . \tag{130}$$

by taking $P_1 = P_2 = P$, (130) reduces to

$$\beta^* = \frac{2}{2^{2(C_{12}+C_{21})} + 1} \; . \tag{131}$$

Therefore the optimal correlation, ρ^*, at infinite SNR is given by

$$\rho^* = \sqrt{1 - \beta^*} = \sqrt{1 - \frac{2}{2^{2(C_{12}+C_{21})} + 1}} \; . \tag{132}$$

The value of ρ^*, for each value of the cooperation link capacities C_{12} and C_{21}, is represented by the horizontal

dashed blue line in the plots shown in Fig. 7. Note that the numerical calculations indeed meet the asymptotic results for large values of SNR.

To conclude, we interpret the numerical and analytical results in terms of the optimal transmission strategies of the users for each SNR regime. Recall that the symbols of the codewords transmitted by the users are modeled by the RVs X_1 and X_2. The fact that for low SNR the correlation is at its maximal value of unity implies that both users tend to transmit the same codewords; this, in turn, indicates that they transmit the same message. However, the only common information the users share is the common message that they have created using the conference. Therefore we conclude that when the channel quality is low, the best strategy for the users is to transmit only the common message and forfeit their private messages (i.e., the parts of their original messages which they have not managed to share). As the SNR grows beyond $\mathrm{SNR}^{\mathrm{Crit}}$, the correlation between the code symbols decreases to some positive value $\rho^* \in (0,1)$, asymptotically approaching (132). This is since when a higher quality channel is experienced, each user transmits not only the common (correlated) message but also his private (uncorrelated) message.

Alternatively, by inspecting the tending of the optimal correlation from the rate perspective some additional intuition unveils. As long as the transmission sum-rate admits $R_1 + R_2 \leq C_{12} + C_{21}$, the transmissions consist only of the correlated common message; namely, the users are fully cooperative. Once the sum-rate surpasses the sum of communication links, i.e. $R_1 + R_2 > C_{12} + C_{21}$, the transmitted code symbols integrate both common and private messages which causes the optimal correlation to drop.

VI. Summary and Concluding Remarks

In this paper we have considered the FSM-MAC with partially cooperative encoders and delayed CSI and derived its capacity region. The achievability proof used another result of this paper, namely, the capacity region of the FSM-MAC with a common message and delayed CSI. The latter result was obtained using rate splitting, multiplexing and simultaneous decoding. This approach circumvents the need to rely on the capacity region's corner points, which becomes cumbersome when their number is large. Furthermore, it is easily extendable and, therefore, establishes the base for simultaneous decoding schemes involving multiple users.

The general conferencing result was then applied to the special case of the Gaussian vector MAC with diagonal channel transfer matrices, which models OFDM-based communication systems. The capacity region is presented in the form of a convex optimization problem and its establishes the optimality of Gaussian Markovian inputs. This result serves as a generalization of [37] to the vector state-dependant case. Focusing on a two-state Gaussian FSM-MAC example, the crucial role of cooperation for low SNR values was demonstrated.

Extensions of the results for the Gaussian vector FSM-MAC to general MIMO settings (see, e.g., [38]), as well as the ISI channel, are currently investigated.

Appendix A
Proof of the Markov Relation in (25)-(27)

We prove the Markov relation (25)-(27) using the following claims. The equality in (25) follows from the fact that $(M_0, S^{q-d_1-1}) - S_{q-d_1} - S_{q-d_2} - S_q$ and so is $(M_0, S_1^{Q-d_1-1}, Q) - S_{Q-d_1} - S_{Q-d_2} - S_Q$.
To show (26) consider the following relations

$$\begin{aligned}
p(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, u_q, q) &= p(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
&= \sum_{m_1\in\mathcal{M}_1} p(m_1, x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
&= \sum_{m_1\in\mathcal{M}_1} p(m_1|s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q)p(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, m_0, m_1, s^{q-d_1-1}, q) \\
&\overset{(a)}{=} \sum_{m_1\in\mathcal{M}_1} p(m_1|s_{q-d_1}, m_0, s^{q-d_1-1}, q)p(x_{1,q}|s_{q-d_1}, m_0, m_1, s^{q-d_1-1}, q) \\
&= \sum_{m_1\in\mathcal{M}_1} p(m_1, x_{1,q}|s_{q-d_1}, m_0, s^{q-d_1-1}, q) \\
&= p(x_{1,q}|s_{q-d_1}, m_0, s^{q-d_1-1}, q)
\end{aligned} \tag{133}$$

where (a) follows from the fact that M_1 is independent of (M_0, S^n) and the fact that $X_{1,q}$ is a deterministic function of $(M_0, M_1, S_{q-d_1}, S^{q-d_1-1})$. Now, since this is true for all q, and because the auxiliary RV is defined as $U = (M_0, S^{Q-d_1-1}, Q)$, it holds that

$$P(x_1|s, \tilde{s}_1, \tilde{s}_2, u) = P(x_1|\tilde{s}_1, u). \tag{134}$$

Finally, to show (27) we use the following relations

$$\begin{aligned}
p(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, u_q, q) &= p(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
&= \sum_{m_2\in\mathcal{M}_2} p(m_2, x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
&= \sum_{m_2\in\mathcal{M}_2} p(m_2|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q)p(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, m_2, s^{q-d_1-1}, q) \\
&\overset{(a)}{=} \sum_{m_1\in\mathcal{M}_1} p(m_2|s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q)p(x_{2,q}|s_{i-d_1}, s_{q-d_2}, m_0, m_2, s^{q-d_1-1}, q) \\
&= \sum_{m_1\in\mathcal{M}_1} p(m_2, x_{2,q}|s_{q-d_1}, s_{q-d_2}, m_0, s^{i-d_1-1}, q) \\
&= p(x_{2,q}|s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q)
\end{aligned} \tag{135}$$

where (a) follows from the fact that M_2 is independent of (X_1, S^n) given M_0 and the fact that $X_{2,i}$ is independent of $(X_{1,q}, S_q)$ given $(M_0, M_2, S_{q-d_1}, S_{q-d_2}, S^{q-d_1-1})$. Again, since the above holds for every q, and by the definition of the RV U, we conclude that

$$P(x_2|x_1, s, \tilde{s}_1, \tilde{s}_2, u) = P(x_2|\tilde{s}_1, \tilde{s}_2, u).$$

APPENDIX B

ANALYSIS OF THE PROBABILITY OF ERROR IN THE ACHIEVABILITY PROOF OF THEOREM 1

First, we analyze the probability of an encoding length mismatch for the component codewords of the sub-messages $m_{0\tilde{s}_1}$, $m_{1\tilde{s}_1}$ or $m_{2\tilde{s}_1,\tilde{s}_2}$, i.e., $\mathbb{P}(N_{\tilde{s}_1} < n_1(\tilde{s}_1))$ and $\mathbb{P}(N_{\tilde{s}_1,\tilde{s}_2} < n_2(\tilde{s}_1,\tilde{s}_2))$. Since the state process is stationary and ergodic, $\lim_{n\to\infty}\frac{N(\tilde{s}_1)}{n} = P(\tilde{s}_1)$ and $\lim_{n\to\infty}\frac{N(\tilde{s}_1,\tilde{s}_2)}{n} = P(\tilde{s}_1,\tilde{s}_2)$ in probability. Therefore, $\mathbb{P}(N_{\tilde{s}_1} < n_2(\tilde{s}_1)) \to 0$ and $\mathbb{P}(N_{\tilde{s}_1,\tilde{s}_2} < n_2(\tilde{s}_1,\tilde{s}_2)) \to 0$ as $n \to \infty$. Therefore, the probability of a decoding error being induced by a mismatch between the actual number of channel state realizations and the expected lengths used in the encoding stage goes to zero.

Next, we analyze the probability of a decoding error. Without loss of generality let us assume that for a given $\widetilde{S}_1 = \tilde{s}_1$, the $k+2$ sub messages of interest that were sent are $(m_{0\tilde{s}_1}, m_{1\tilde{s}_1}, \mathbf{m}_2(\tilde{s}_1)) = (1,1,\mathbf{1})$, where $\mathbf{1}$ is a k-dimensional vector of 1's. First, we note that an error in the decoding of $X_2^{n_1(\tilde{s}_1)}(\hat{m}_{0\tilde{s}_1}, \hat{\mathbf{m}}_2(\tilde{s}_1))$ can occur as a result of an error in any subset of the component reconstructed sub-messages of $\hat{\mathbf{m}}_2(\tilde{s}_1)$, i.e., in any subset of the set $\{\hat{m}_{2,\tilde{s}_1,1}, \hat{m}_{2,\tilde{s}_1,2}, ..., \hat{m}_{2,\tilde{s}_1,k}\}$. In order to deal with such an event, we define the following two sets:

$$\mathcal{S}_T = \{\tilde{s}_2 \in \mathcal{S} : \hat{m}_{2\tilde{s}_1\tilde{s}_2} = 1\}, \tag{136}$$

$$\mathcal{S}_F = \mathcal{S}_T^C = \{\tilde{s}_2 \in \mathcal{S} : \hat{m}_{2\tilde{s}_1\tilde{s}_2} \neq 1\}. \tag{137}$$

$\mathcal{S}_T$ is the subset that contains all the states $\tilde{s}_2 \in \mathcal{S}$ for which no decoding error occurred and the reconstructed sub-message is correct (here the subscript T stands for 'True'). $\mathcal{S}_F$ is the subset that contains all the states $\tilde{s}_2 \in \mathcal{S}$ for which a decoding error occurred and the reconstructed sub-message is incorrect (here the subscript F stands for 'False').

We note that if, for some $\tilde{s}_1 \in \mathcal{S}$, and error occurred in all reconstructed sub-messages $\hat{m}_{2,\tilde{s}_1,\tilde{s}_2}$, where $\tilde{s}_2 \in \mathcal{S}$, we have $\mathcal{S}_T = \emptyset$ and $\mathcal{S}_F = \mathcal{S}$. If no error occurred we have $\mathcal{S}_F = \emptyset$ and $\mathcal{S}_T = \mathcal{S}$. Moreover, we note that the reconstructed codeword $X_2^{n_1(\tilde{s}_1)}$ can be now described as function of the common sub-message $\hat{m}_{0\tilde{s}_1}$, the two sets (136)-(137) and a set of $|\mathcal{S}_F|$ indices $q_{\tilde{s}_2}$. The latter set states for each sub-message $\hat{m}_{2\tilde{s}_1\tilde{s}_2} \neq 1$, where $\tilde{s}_2 \in \mathcal{S}_F$, which is the appropriate index, $q_{\tilde{s}_2} \in \{2,3,\ldots,2^{n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)}\}$, to which it equals. Therefore, henceforth we use the notation $X_2^{n_1(\tilde{s}_1)}(\hat{m}_{0\tilde{s}_1}, \mathbf{v}_{\mathcal{S}_F})$, where $\mathbf{v}_{\mathcal{S}_F} = (v(1), v(2), \ldots, v(k))$ is a k-dimensional vector of indices such that

$$v(\tilde{s}_2) = \left\{ \begin{array}{cc} 1, & \tilde{s}_2 \in \mathcal{S}_T \\ q_{\tilde{s}_2}, & \tilde{s}_2 \in \mathcal{S}_F \end{array} \right\}. \tag{138}$$

Note that for every $\mathcal{S}_F \neq \emptyset$ we have $\mathbf{v}_{\mathcal{S}_F} \neq \mathbf{1}$.
In addition to these definitions we introduce the following lemma.

Lemma 6 *Let $P_{XY}(x,y)$ denote the joint distribution of two RV's (X,Y) on $\mathcal{X}\times\mathcal{Y}$, let $P_X(x)$ and $P_Y(y)$ denote their marginal distributions and let $n = n_1 + n_2$. Further, let the sequences (x^{n_1}, y^{n_1}) and (x^{n_2}, y^{n_2}) be drawn in an i.i.d manner according to $P_{XY}(x,y)$ and $P_X(x)P_Y(y)$, respectively. Let (x^n, y^n) be a concatenation of*

(x^{n_1}, y^{n_1}) and (x^{n_2}, y^{n_2}). Then,

$$(1-\delta'_{\epsilon,n}) \cdot 2^{-n_2(I(X;Y)-\delta_\epsilon)} \le \mathbb{P}\big\{(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}\big\} \le 2^{-n_2(I(X;Y)+\delta_\epsilon)}, \tag{139}$$

where $\delta_\epsilon \to 0$ and $\delta'_{\epsilon,n} \to 0$ as $\epsilon \to 0$.

The proof of Lemma 6 is given in Appendix E. Lemma 6 gives rise to another observation.

Lemma 7 *Let $P_{XYWU}(x,y,u,w)$ denote the joint distribution of the tuple of RVs (X,Y,U,W) on $\mathcal{X}\times\mathcal{Y}\times\mathcal{U}\times\mathcal{W}$. Let $P_{XY|WU}(x,y|w,u)$ denote the conditional joint distribution of X,Y given U,W, let $P_{X|WU}(x|w,u)$ and $P_{Y|WU}(y|w,u)$ denote the corresponding marginal conditional distributions of X and Y. Define $n_w = P_W(W = w)\cdot n$, clearly $n = \sum_{w\in\mathcal{W}} n_w$. Let $\mathcal{W}_F = \{w \in \mathcal{W} : P_{XY|UW=w} = P_{X|UW=w}P_{Y|UW=w}\}$. Let $\big\{(x^{n_w}, y^{n_w})\big\}_{w\in\mathcal{W}}$ be a set of sequences which are drawn in an i.i.d manner according to $P_{X|UW=w}P_{Y|UW=w}$ when $w \in \mathcal{W}_F$, and according to $P_{XY|UW=w}$ when $w \in \mathcal{W}_F^C$. Let (x^n, y^n) be a concatenation of the set of sequences $\big\{(x^{n_w}, y^{n_w})\big\}_{w\in\mathcal{W}}$. Then,*

$$(1-\delta'_{\epsilon,n}) \cdot 2^{-\sum_{w\in\mathcal{W}_F} n_w(I(X;Y|U,W=w)-\delta_\epsilon)} \le \mathbb{P}\big\{(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}\big\} \le 2^{-\sum_{w\in\mathcal{W}_F} n_w(I(X;Y|U,W=w)+\delta_\epsilon)}, \tag{140}$$

where $\delta_\epsilon \to 0$ and $\delta'_{\epsilon,n} \to 0$ as $\epsilon \to 0$.

Lemma 7 follows directly from Lemma 6 and its proof is therefore omitted. Note that even though both lemmas refer to sequences (x^n, y^n) that are created by concatenation, the results can be generalized to sequences (x^n, y^n) that are created by any mixture of the sub-sequences (and not necessarily concatenation).

Next, for simplicity, we use the notation in (33)-(38) and introduce the following

$$n_1(\tilde{s}_1) = n_1, \tag{141}$$

$$n_2(\tilde{s}_1, \tilde{s}_2 = \ell) = n_{2\ell}. \tag{142}$$

Note that,

$$n_{2\ell} = n_1 P_\ell, \tag{143}$$

for a proper choice of ϵ' and ϵ'' (namely, $\epsilon'' = \epsilon' P_\ell$).

Using these notations we can define the events that correspond to all possible decoding errors (recall (28)). We start with the events in which no errors occurred in the decoding of the sub-messages of $\mathbf{m}_2(\tilde{s}_1)$ (i.e., in these events $\hat{\mathbf{m}}_2(\tilde{s}_1)$ is presumed to be correct):

$$E_1 = \big\{(U^{n_1}(1), X_1^{n_1}(1,1), X_2^{n_1}(1,\mathbf{1}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \notin \mathcal{T}_\epsilon^{(n_1)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}, \tag{144}$$

$$E_2 = \big\{\exists i \neq 1 : (U^{n_1}(i), X_1^{n_1}(i,1), X_2^{n_1}(i,\mathbf{1}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \in \mathcal{T}_\epsilon^{(n_1)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}, \tag{145}$$

$$E_3 = \big\{\exists j \neq 1 : (U^{n_1}(1), X_1^{n_1}(1,j), X_2^{n_1}(1,\mathbf{1}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \in \mathcal{T}_\epsilon^{(n_1)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}, \tag{146}$$

$$E_4 = \big\{\exists (i,j) \neq (1,1) : (U^{n_1}(i), X_1^{n_1}(i,j), X_2^{n_1}(i,\mathbf{1}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \in \mathcal{T}_\epsilon^{(n_1)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}. \tag{147}$$

Next, we present the events in which $\hat{\mathbf{m}}_2(\tilde{s}_1)$ is fully or partly incorrect. For a given set $\mathcal{S}_F \subseteq \mathcal{S}$, let us define:

$$E_5(\mathcal{S}_F) = \big\{\exists \{q_{\tilde{s}_2}\}_{\tilde{s}_2 \in \mathcal{S}_F}, \mathbf{v}_{\mathcal{S}_F} \neq \mathbf{1} : (U^{n_1}(1), X_1^{n_1}(1,1), X_2^{n_1}(1, \mathbf{v}_{\mathcal{S}_F}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \in \mathcal{T}_\epsilon^{(n)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}, \quad (148)$$

$$E_6(\mathcal{S}_F) = \big\{\exists j \neq 1, \exists \{q_{\tilde{s}_2}\}_{\tilde{s}_2 \in \mathcal{S}_F}, \mathbf{v}_{\mathcal{S}_F} \neq \mathbf{1} : (U^{n_1}(1), X_1^{n_1}(1,j), X_2^{n_1}(1, \mathbf{v}_{\mathcal{S}_F}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \in \mathcal{T}_\epsilon^{(n)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}, \quad (149)$$

$$E_7(\mathcal{S}_F) = \big\{\exists i \neq 1, \exists \{q_{\tilde{s}_2}\}_{\tilde{s}_2 \in \mathcal{S}_F}, \mathbf{v}_{\mathcal{S}_F} \neq \mathbf{1} : (U^{n_1}(i), X_1^{n_1}(i,1), X_2^{n_1}(i, \mathbf{v}_{\mathcal{S}_F}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \in \mathcal{T}_\epsilon^{(n)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}, \quad (150)$$

$$E_8(\mathcal{S}_F) = \big\{\exists (i,j) \neq (1,1), \exists \{q_{\tilde{s}_2}\}_{\tilde{s}_2 \in \mathcal{S}_F}, \mathbf{v}_{\mathcal{S}_F} \neq \mathbf{1} : (U^{n_1}(i), X_1^{n_1}(i,j), X_2^{n_1}(i, \mathbf{v}_{\mathcal{S}_F}), Y^{n_1}, S^{n_1}, \widetilde{S}_2^{n_1}) \in \mathcal{T}_\epsilon^{(n)} \big| \widetilde{S}_1 = \tilde{s}_1\big\}. \quad (151)$$

Using the union of events bound, we can bound the probability of a decoding error by:

$$P_e = \mathbb{P}\bigg(\Big\{\bigcup_{i=1}^{4} E_i\Big\} \bigcup \Big\{\bigcup_{\mathcal{S}_F \subseteq \mathcal{S}} \bigcup_{i=5}^{8} E_i(\mathcal{S}_F)\Big\}\bigg) \leq \sum_{i=1}^{4} \mathbb{P}(E_i) + \sum_{\mathcal{S}_F \subseteq \mathcal{S}} \sum_{i=5}^{8} \mathbb{P}\big(E_i(\mathcal{S}_F)\big) \quad (152)$$

We need to show that for the coding scheme presented in Section III-B, and for a rate triplet (R_0, R_2, R_2) as given in Theorem 1, $P_e \to 0$ as $n \to \infty$.

1) $\mathbb{P}(E_1) \to 0$ as $n_1 \to \infty$ from the law of large numbers.

2) In order to upper bound $\mathbb{P}(E_3)$ consider:

$$\begin{aligned}
\mathbb{P}(E_3) &= \sum_{j=2}^{2^{n_1 R_1'}} \mathbb{P}(E_{3,j}) \\
&\leq \sum_{j=2}^{2^{n_1 R_1'}} 2^{-n_1\left(I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2)+\delta_\epsilon\right)} \\
&\leq 2^{n_1\left(R_1' - I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2)+\delta_\epsilon\right)}
\end{aligned}$$

So in order to have $\mathbb{P}(E_3) \to 0$ as $n_1 \to \infty$ the following must hold:

$$R_1' < I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2). \quad (153)$$

We proceed with the upper bounding of (148), (149) and (151), for a given set $\mathcal{S}_F \subseteq \mathcal{S}$.

3) In order to upper bound $\mathbb{P}\big(E_5(\mathcal{S}_F)\big)$ for a given set $\mathcal{S}_F$, first consider:

$$\begin{aligned}
\mathbb{P}\big(E_5(\mathcal{S}_F)\big) &\overset{(a)}{\leq} 2^{-\sum_{\ell \in \mathcal{S}_F} n_{2\ell}\left(I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\delta_\epsilon\right)} \cdot \prod_{\ell \in \mathcal{S}_F} 2^{n_{2\ell} R_{2\ell}} \\
&= 2^{-\sum_{\ell \in \mathcal{S}_F} n_{2\ell}\left(R_{2\ell} - I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\delta_\epsilon\right)}
\end{aligned}$$

$$= 2^{-n_1 \sum_{\ell \in \mathcal{S}_F} P_\ell \left(R_{2\ell} - I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell) + \delta_\epsilon \right)} \tag{154}$$

where (a) follows by applying the union bound and Lemma 7. This is because the errors of interest occur in the component codewords $X_2^{n_{2\ell}}(1, q_\ell)$, for every $q_\ell \in \{2, 3, \ldots, 2^{n_{2\ell} R_{2\ell}}\}$, where $\ell \in \mathcal{S}_F$. According to Lemma 7, the probability of each of those errors is upper bounded by $2^{-n_{2\ell}\left(I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\delta_\epsilon\right)}$. Note that here the codewords $X_1^{n_1}(1)$ and $U^{n_1}(1)$ are fixed and correct for all $\mathcal{S}_F \subseteq \mathcal{S}$, which explains the structure of the mutual information term in (154). Therefore, in order to have $\mathbb{P}\big(E_5(\mathcal{S}_F)\big) \to 0$ as $n_1 \to \infty$, the following must hold,

$$\sum_{\ell \in \mathcal{S}_F} P_\ell R_{2\ell} < \sum_{\ell \in \mathcal{S}_F} P_\ell I(X_2; Y | X_1, U, S, \widetilde{S}_1 = \tilde{s}_1, \widetilde{S}_2 = \ell), \tag{155}$$

for every possible set $\mathcal{S}_F \subseteq \mathcal{S}$.

4) In order to upper bound $\mathbb{P}\big(E_6(\mathcal{S}_F)\big)$ for a given set $\mathcal{S}_F$, we have

$$\begin{aligned}
\mathbb{P}\big(E_6(\mathcal{S}_F)\big) &\overset{(a)}{\leq} 2^{-\sum_{\ell \in \mathcal{S}_F} n_{2\ell}\left(I(X_1,X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\delta_\epsilon\right)} \\
&\quad \cdot 2^{-\sum_{t \in \mathcal{S}_T} n_{2t}\left(I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=t)+\delta_\epsilon\right)} \cdot 2^{n_1 R_1'} \cdot \prod_{\ell \in \mathcal{S}_F} 2^{n_{2\ell} R_{2\ell}} \\
&= 2^{n_1 R_1' + \sum_{\ell \in \mathcal{S}_F} n_{2\ell} R_{2\ell} - \sum_{\ell \in \mathcal{S}_F} n_{2\ell}\left(I(X_1,X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\delta_\epsilon\right)} \\
&\quad \cdot 2^{-\sum_{t \in \mathcal{S}_T} n_{2t}\left(I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=t)+\delta_\epsilon\right)} \\
&= 2^{n_1\left(R_1' + \sum_{\ell \in \mathcal{S}_F} P_\ell\left(R_{2\ell} - I(X_1,X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell) - \delta_\epsilon\right) - \sum_{t \in \mathcal{S}_T} P_t\left(I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=t)+\delta_\epsilon\right)\right.}
\end{aligned} \tag{156}$$

where (a) follows by applying the union bound and Lemma 7. First, note that if the component codeword $X_1^{n_1}(1, j)$, where $j \in \{2, 3, \ldots, 2^{n_1 R_1'}\}$, is incorrect, then all the sub-component codewords $X_1^{n_{2\ell}}(1, j)$, for every $\ell \in \mathcal{S}$ (and in particular for every $\ell \in \mathcal{S}_F$), are also incorrect. Moreover, errors also occur in the component codewords $X_2^{n_{2\ell}}(1, q_\ell)$ for every $q_\ell \in \{2, 3, \ldots, 2^{n_{2\ell} R_{2\ell}}\}$ and $\ell \in \mathcal{S}_F$. According to Lemma 7, the probability of each of the errors for $\ell \in \mathcal{S}_F$ is upper bounded by $2^{-n_{2\ell}\left(I(X_1,X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\delta_\epsilon\right)}$ (here both $X_1^{n_{2\ell}}$ and $X_2^{n_{2\ell}}$ are incorrect), whereas the probability of each of the errors for $t \in \mathcal{S}_T$ is upper bounded by $2^{-n_{2t}\left(I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=t)+\delta_\epsilon\right)}$ (here only $X_1^{n_{2t}}$ is incorrect). Finally, note that the codeword $U^{n_1}(1)$ is fixed and correct for all $\mathcal{S}_F \subseteq \mathcal{S}$, which explains the structure of the mutual information terms in (156). Therefore, in order to have $\mathbb{P}\big(E_6(\mathcal{S}_F)\big) \to 0$ as $n_1 \to \infty$, the following must hold,

$$R_1' + \sum_{\ell \in \mathcal{S}_F} P_\ell R_{2\ell} < \sum_{\ell \in \mathcal{S}_F} P_\ell I(X_1, X_2; Y | U, S, \widetilde{S}_1 = \tilde{s}_1, \widetilde{S}_2 = \ell) + \sum_{t \in \mathcal{S}_T} P_t I(X_1; Y | X_2, U, S, \widetilde{S}_1 = \tilde{s}_1, \widetilde{S}_2 = t), \tag{157}$$

for every possible set $\mathcal{S}_F \subseteq \mathcal{S}$.

The inequality in (157) can also be rewritten as,

$$\begin{aligned}R_1' + \sum_{\ell\in\mathcal{S}_F} P_\ell R_{2\ell} &< \sum_{\ell\in\mathcal{S}_F} P_\ell\big(I(X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)\big)\\ &\quad+\sum_{t\in\mathcal{S}_T} P_t I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=t)\\ &=\sum_{\ell\in\mathcal{S}_F} P_\ell I(X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\sum_{t\in\mathcal{S}} P_t I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=t)\\ &=I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2)+\sum_{\ell\in\mathcal{S}_F} P_\ell I(X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)\end{aligned} \tag{158}$$

5) In order to upper bound $\mathbb{P}\big(E_8(\mathcal{S}_F)\big)$ for a given set $\mathcal{S}_F$, first consider:

$$\begin{aligned}\mathbb{P}\big(E_8(\mathcal{S}_F)\big) &\overset{(a)}{\leq} 2^{-\sum_{\ell\in\mathcal{S}} n_{2\ell}\left(I(U,X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell)+\delta_\epsilon\right)}\cdot 2^{n_1R_0'}\cdot 2^{n_1R_1'}\cdot\prod_{\ell\in\mathcal{S}} 2^{n_{2\ell}R_{2\ell}}\\ &=2^{n_1\left(R_0'+R_1'+R_2'-I(U,X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2)+\delta_\epsilon\right)}\end{aligned}$$

where (a) follows from the union bound and from the fact that there is an error in the component codeword $U^{n_1}(i)$, where $i\in\{2,3,\ldots,2^{n_l R_0'}\}$. This is because an incorrect common sub-message $\hat{m}_{0\tilde{s}_1}\neq 1$, for $\tilde{s}_1\in\mathcal{S}$, causes an error in all other component codewords. Namely, errors occur in the component codeword $X_1^{n_1}(i,j)$, for every $j\in\{2,3,\ldots,2^{n_1R_1'}\}$, and in each of the component codewords $X_2^{n_{2\ell}}(i,q_\ell)$, for every $q_\ell\in\{2,3,\ldots,2^{n_{2\ell}R_{2\ell}}\}$ when $\ell\in\mathcal{S}_F$, and for $q_\ell=1$ when $\ell\in\mathcal{S}_T$. In other words, an error in the common sub-message will cause all three codewords to be incorrect for every $\ell\in\mathcal{S}$. Therefore, the probability of error for the whole block (of length n_1) is upper bounded by $2^{n_1\left(I(U,X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2)+\delta_\epsilon\right)}$, and in order to have $\mathbb{P}\big(E_8(\mathcal{S}_F)\big)\to 0$ as $n\to\infty$, the following must hold,

$$R_0'+R_1'+R_2'<I(U,X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{159}$$

for every possible set $\mathcal{S}_F\subseteq\mathcal{S}$. We rewrite the inequality in (159) as,

$$R_0'+R_1'+R_2'\overset{(a)}{<}I(X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2)+I(U;Y|X_1,X_2,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2) \tag{160}$$

$$\overset{(b)}{=}I(X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{161}$$

where (a) follows from the mutual information chain rule and (b) follows from the fact that Y is independent of U given (X_1,X_2,S), by the underlying channel model (see Subsection II-A).

Note that the restrictions that arise from upper bounding the probability of the events in (145), (147) and (150) are all redundant given the restriction that is obtained by upper bounding the probability of the event in (151) presented above. This is because in all these events the codeword U^{n_1} is incorrect, which causes the codewords $X_1^{n_1}$ and $X_2^{n_1}$ to be incorrect as well. Thus, the restrictions that must hold in order for these probabilities to go to zero contain the *same* mutual information term as an upper bound on the

corresponding rates. However, upper bounding the probability of (151) yields a restriction on $R_0+R_1+R_2$, whereas upper bounding the probabilities of (145), (147) and (150) yields restrictions on R_0, R_0+R_1 and R_0+R_2, respectively. Due to the fact that $R_j \geq 0$ for all $j \in \{0,1,2\}$, the three latter restrictions become redundant. For this reason the upper bounding on the probabilities of (145), (147) and (150) is omitted.

Summarizing the above results, we get that the probability of error, conditioned on a particular codeword being sent, goes to zero if the following conditions are met:

$$R_1' < I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{162}$$

$$\sum_{\ell\in\mathcal{S}_F} P_\ell R_{2\ell} < \sum_{\ell\in\mathcal{S}_F} P_\ell I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell), \tag{163}$$

$$R_1' + \sum_{\ell\in\mathcal{S}_F} P_\ell R_{2\ell} < I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2) + \sum_{\ell\in\mathcal{S}_F} P_\ell I(X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell), \tag{164}$$

$$R_0' + R_1' + R_2' < I(X_1,X_2;Y|S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{165}$$

for every possible set $\mathcal{S}_F \subseteq \mathcal{S}$.

APPENDIX C
FOURIER-MOTZKIN ELIMINATION FOR THE PROOF OF THEOREM 1

In this appendix we use the FME to show that the set of inequalities:

$$\sum_{\ell\in\mathcal{S}_F} P_\ell R_{2\ell} < \sum_{\ell\in\mathcal{S}_F} P_\ell I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell), \tag{166}$$

$$R_1' + \sum_{\ell\in\mathcal{S}_F} P_\ell R_{2\ell} < I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2) + \sum_{\ell\in\mathcal{S}_F} P_\ell I(X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell), \tag{167}$$

for all $\mathcal{S}_F \subseteq \mathcal{S}$, are equivalent to the two inequalities,

$$R_2' < I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2), \tag{168}$$

$$R_1' + R_2' < I(X_1,X_2;Y|U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2). \tag{169}$$

Recall that the state space is $\mathcal{S} = \{1,2,\dots,k\}$. For simplicity, we henceforth denote an arbitrary subset of $\mathcal{S}$ by $\mathcal{A}$ rather than $\mathcal{S}_F$ as in (166)-(167). Denote by $\mathcal{S}_\mathcal{A}$ the compliment of $\mathcal{A}$ in $\mathcal{S}$, that is $\mathcal{S}_\mathcal{A} = \mathcal{A}^C$ (where the compliment is taken with respect to the state space $\mathcal{S}$). Specifically, $\mathcal{S}_{\ell_1,\ell_2,\dots,\ell_p} = \mathcal{S}\backslash\{\ell_1,\ell_2,\dots,\ell_p\}$ where $\ell_i \in \mathcal{S}$ for all $i \in \{1,2,\dots,p\}$ and $p \leq k$. Moreover, throughout this appendix we use the notations in (33)-(38) and (141)-(142). Accordingly, we have:

$$R_2' = \sum_{\ell\in\mathcal{S}} P_\ell R_{2\ell}. \tag{170}$$

In addition, since all involved mutual information terms are conditional on $(U, S, \widetilde{S}_1 = \tilde{s}_1)$, we use the following shortened notation:

$$\begin{aligned} I(X_1;Y|X_2,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2) &\triangleq \mathcal{I}(X_1;Y|X_2), \\ I(X_2;Y|X_1,U,S,\widetilde{S}_1=\tilde{s}_1,\widetilde{S}_2=\ell) &\triangleq \mathcal{I}_\ell(X_2;Y|X_1), \end{aligned} \tag{171}$$

For clarity, we split the FME process into two stages where the second stage serves as an inductive stage that concludes the proof.

Stage 1: In this stage we eliminate the partial rate R_{21} (and all its associated components) from the set of inequalities (166)-(167) in order to remain with inequalities for subsets $\mathcal{A} \subseteq \mathcal{S}_1$ only.

First, consider the inequality (166) for an arbitrary subset $\mathcal{B} \subseteq \mathcal{S}$ that contains the element $\{1\}$. The subset $\mathcal{B}$ can be represented by $\mathcal{B} = \mathcal{A} \bigcup \{1\}$ for some $\mathcal{A} \subseteq \mathcal{S}_1$. Using (170), we rewrite the left hand side (LHS) of this inequality as follows:

$$\sum_{\ell\in\mathcal{B}} P_\ell R_{2\ell} = R_2' - \sum_{\ell\in\mathcal{B}^C} P_\ell R_{2\ell}. \tag{172}$$

Using similar arguments, for the right hand side (RHS) of (166), we have that (see (171)):

$$\sum_{\ell\in\mathcal{B}} P_\ell \mathcal{I}_\ell(X_2;Y|X_1) = \mathcal{I}(X_2;Y|X_1) - \sum_{\ell\in\mathcal{B}^C} P_\ell \mathcal{I}_\ell(X_2;Y|X_1). \tag{173}$$

However, $\mathcal{B}^C \subseteq \mathcal{S}_1$, thus R_{21} is eliminated from all inequalities of the form (166).

A similar procedure can be applied to the inequalities of the form (167), taken for subsets $\mathcal{B} \subseteq \mathcal{S}$ containing the element $\{1\}$. The LHS is rewritten using (170), whereas for the RHS consider:

$$\begin{aligned} \mathcal{I}(X_1;Y|X_2) + \sum_{\ell\in\mathcal{B}} P_\ell \mathcal{I}_\ell(X_2;Y) &= \mathcal{I}(X_1;Y|X_2) + \mathcal{I}(X_2;Y) - \sum_{\ell\in\mathcal{B}^C} P_\ell \mathcal{I}_\ell(X_2;Y) \\ &= \mathcal{I}(X_1,X_2;Y) - \sum_{\ell\in\mathcal{B}^C} P_\ell \mathcal{I}_\ell(X_2;Y). \end{aligned} \tag{174}$$

By doing so R_{21} is eliminated from all inequalities of the form (167), and thus we are left with inequalities of four different forms:

$$\sum_{\ell\in\mathcal{A}} P_\ell R_{2\ell} < \sum_{\ell\in\mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y|X_1), \tag{175}$$

$$R_2' - \sum_{\ell\in\mathcal{A}} P_\ell R_{2\ell} < \mathcal{I}(X_2;Y|X_1) - \sum_{\ell\in\mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y|X_1), \tag{176}$$

$$R_1' + \sum_{\ell\in\mathcal{A}} P_\ell R_{2\ell} < \mathcal{I}(X_1;Y|X_2) + \sum_{\ell\in\mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y), \tag{177}$$

$$R_1' + R_2' - \sum_{\ell\in\mathcal{A}} P_\ell R_{2\ell} < \mathcal{I}(X_1,X_2;Y) - \sum_{\ell\in\mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y), \tag{178}$$

for every $\mathcal{A} \subseteq \mathcal{S}_1$.

Stage 2: Next, we eliminate the partial rate R_{22} (and all its associated components) from the set of inequalities (175)-(178) in order to remain with inequalities for subsets $\mathcal{A} \subseteq \mathcal{S}_{1,2}$ only. In order to do so, we divide the remaining inequalities into two classes. The first contains all the inequalities of the forms (175)-(178) taken for subsets $\mathcal{A} \subseteq \mathcal{S}_{1,2}$, that is subsets $\mathcal{A}$ that do not contain both elements $\{1,2\}$. The second class consists of all inequalities for which the subset contains the element $\{2\}$. We denote such a subset by $\mathcal{B} \subseteq \mathcal{S}_1$ (with some abuse of notation), where $\mathcal{B} = \mathcal{A} \bigcup \{2\}$ for some $\mathcal{A} \subseteq \mathcal{S}_{1,2}$. Using the FME we eliminate the partial rate R_{22} and all its associated components from these inequalities. By doing so, new restrictions arise. However, we show that these restrictions are redundant as they are all admitted by having the inequalities of the first class (namely, the inequalities of the forms (175)-(178) taken for subsets $\mathcal{A} \subseteq \mathcal{S}_{1,2}$).

By separating the components which involve the element $\{2\}$ in the second class of inequalities, the latter can be rewritten as:

$$P_2 R_{22} + \sum_{\ell \in \mathcal{A}} P_\ell R_{2\ell} < P_2 \mathcal{I}_2(X_2;Y|X_1) + \sum_{\ell \in \mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y|X_1), \tag{179}$$

$$R_2' - P_2 R_{22} - \sum_{\ell \in \mathcal{A}} P_\ell R_{2\ell} < \mathcal{I}(X_2;Y|X_1) - P_2 \mathcal{I}_2(X_2;Y|X_1) - \sum_{\ell \in \mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y|X_1), \tag{180}$$

$$R_1' + P_2 R_{22} + \sum_{\ell \in \mathcal{A}} P_\ell R_{2\ell} < \mathcal{I}(X_1;Y|X_2) + P_2 \mathcal{I}_2(X_2;Y) + \sum_{\ell \in \mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y), \tag{181}$$

$$R_1' + R_2' - P_2 R_{22} - \sum_{\ell \in \mathcal{A}} P_\ell R_{2\ell} < \mathcal{I}(X_1,X_2;Y) - P_2 \mathcal{I}_2(X_2;Y) - \sum_{\ell \in \mathcal{A}} P_\ell \mathcal{I}_\ell(X_2;Y), \tag{182}$$

for every $\mathcal{A} \subseteq \mathcal{S}_{1,2}$ (since $\mathcal{B} \backslash \{2\} = \mathcal{A}$, for some $\mathcal{A} \subseteq \mathcal{S}_{1,2}$).

For the inequalities of the forms (179)-(182), we apply the FME through the following steps:

1) Combining (179)-(180) taken for some subsets $\mathcal{A}^{'}, \mathcal{A}^{''} \subseteq \mathcal{S}_{1,2}$, respectively, we get:

$$R_2' - \sum_{\ell \in \mathcal{A}''} P_\ell R_{2\ell} + \sum_{\ell \in \mathcal{A}'} P_\ell R_{2\ell} < \mathcal{I}(X_2;Y|X_1) - \sum_{\ell \in \mathcal{A}''} P_\ell \mathcal{I}_\ell(X_2;Y|X_1) + \sum_{\ell \in \mathcal{A}'} P_\ell \mathcal{I}_\ell(X_2;Y|X_1). \tag{183}$$

But the above inequality can be constructed by adding (175) taken for the subset $\mathcal{A}^{'}$ to (176) taken for the subset $\mathcal{A}^{''}$. We thus conclude that the restriction in (183) is redundant.

2) Next, combining (181)-(182) taken for some subsets $\mathcal{A}^{'}, \mathcal{A}^{''} \subseteq \mathcal{S}_{1,2}$, respectively, we get:

$$\begin{aligned} R_1' + \sum_{\ell \in \mathcal{A}'} P_\ell R_{2\ell} &+ (R_1' + R_2') - \sum_{\ell \in \mathcal{A}''} P_\ell R_{2\ell} \\ &< \mathcal{I}(X_1,X_2;Y) + \sum_{\ell \in \mathcal{A}'} P_\ell \mathcal{I}_\ell(X_2;Y) + \mathcal{I}(X_1;Y|X_2) - \sum_{\ell \in \mathcal{A}''} P_\ell \mathcal{I}_\ell(X_2;Y). \end{aligned} \tag{184}$$

The above inequality can be derived by adding (177) taken for the subset $\mathcal{A}^{'}$ to (178) taken for the subset $\mathcal{A}^{''}$, and it is thus also redundant.

3) By combining (179) and (182) taken for some subsets $\mathcal{A}^{'}, \mathcal{A}^{''} \subseteq \mathcal{S}_{1,2}$, respectively, we get:

$$\begin{aligned}
&\sum_{\ell\in\mathcal{A}'} P_\ell R_{2\ell} + R_1' + R_2' - \sum_{\ell\in\mathcal{A}''} P_\ell R_{2\ell} \\
&\quad < P_2\mathcal{I}_2(X_2;Y|X_1) + \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1) + \mathcal{I}(X_1,X_2;Y) - P_2\mathcal{I}_2(X_2;Y) - \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y) \\
&\quad \overset{(a)}{=} P_2\big(\mathcal{I}_2(X_2;Y,X_1) - \mathcal{I}_2(X_2;Y)\big) + \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1) + \mathcal{I}(X_1,X_2;Y) - \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y) \\
&\quad = P_2\mathcal{I}_2(X_2;X_1|Y) + \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1) + \mathcal{I}(X_1,X_2;Y) - \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y) \qquad (185)
\end{aligned}$$

where (a) follows from the fact that X_1 is independent of X_2 given $(U, \widetilde{S}_1, \widetilde{S}_2)$. Note that by adding inequality (175) taken for the subset $\mathcal{A}^{'}$ to inequality (178) taken for the subset $\mathcal{A}^{''}$, we get:

$$\sum_{\ell\in\mathcal{A}'} P_\ell R_{2\ell} + R_1' + R_2' - \sum_{\ell\in\mathcal{A}''} P_\ell R_{2\ell} < \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1) + \mathcal{I}(X_1,X_2;Y) - \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y),$$

which is tighter than (185) due to the non-negativity of the mutual information. This implies that (185) is redundant as well.

4) Finally, by combining (180) and (181) taken for some subsets $\mathcal{A}^{'}, \mathcal{A}^{''} \subseteq \mathcal{S}_{1,2}$, respectively, we get:

$$\begin{aligned}
&R_2' - \sum_{\ell\in\mathcal{A}'} P_\ell R_{2\ell} + R_1' + \sum_{\ell\in\mathcal{A}''} P_\ell R_{2\ell} \\
&\qquad < \mathcal{I}(X_2;Y|X_1) - P_2\mathcal{I}_2(X_2;Y|X_1) - \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1) \\
&\qquad + \mathcal{I}(X_1;Y|X_2) + P_2\mathcal{I}_2(X_2;Y) + \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y). \qquad (186)
\end{aligned}$$

Denoting $\widetilde{\mathcal{A}}^{\,'} \triangleq \mathcal{A}^{'} \Big\backslash \Big\{\mathcal{A}^{'} \bigcap \mathcal{A}^{''}\Big\}$ and $\widetilde{\mathcal{A}}^{\,''} \triangleq \mathcal{A}^{''} \Big\backslash \Big\{\mathcal{A}^{'} \bigcap \mathcal{A}^{''}\Big\}$ we rewrite (186) as,

$$\begin{aligned}
&R_2' - \sum_{\ell\in\widetilde{\mathcal{A}}'} P_\ell R_{2\ell} + R_1' + \sum_{\ell\in\widetilde{\mathcal{A}}''} P_\ell R_{2\ell} \\
&\quad < \mathcal{I}(X_2;Y|X_1) - P_2\mathcal{I}_2(X_2;Y|X_1) - \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1) + \mathcal{I}(X_1;Y|X_2) + P_2\mathcal{I}_2(X_2;Y) \\
&\quad + \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y) \\
&\quad \overset{(a)}{=} \mathcal{I}(X_1;Y|X_2) + \mathcal{I}(X_2;Y,X_1) + P_2\Big(\mathcal{I}_2(X_2;Y) - \mathcal{I}_2(X_2;Y,X_1)\Big) + \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y) \\
&\quad - \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1) \\
&\quad = \mathcal{I}(X_1;Y|X_2) + \mathcal{I}(X_2;Y) + \mathcal{I}(X_2;X_1|Y) + P_2\Big(\mathcal{I}_2(X_2;Y) - \mathcal{I}_2(X_2;Y) - \mathcal{I}_2(X_2;X_1|Y)\Big) \\
&\quad + \sum_{\ell\in\mathcal{A}''} P_\ell\mathcal{I}_\ell(X_2;Y) - \sum_{\ell\in\mathcal{A}'} P_\ell\mathcal{I}_\ell(X_2;Y|X_1)
\end{aligned}$$

$$= \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) - P_2 \mathcal{I}_2(X_2; X_1|Y) + \sum_{\ell \in \mathcal{A}''} P_\ell \mathcal{I}_\ell(X_2; Y)$$

$$- \sum_{\ell \in \mathcal{A}'} P_\ell \mathcal{I}_\ell(X_2; Y|X_1)$$

$$= \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}\setminus\{2\}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) + \sum_{\ell \in \mathcal{A}''} P_\ell \mathcal{I}_\ell(X_2; Y) - \sum_{\ell \in \mathcal{A}'} P_\ell \mathcal{I}_\ell(X_2; Y|X_1)$$

$$\stackrel{(b)}{=} \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}\setminus\left\{\mathcal{A}'' \bigcup \{2\}\right\}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) + \sum_{\ell \in \mathcal{A}''} P_\ell \Big(\mathcal{I}_\ell(X_2; X_1|Y) + \mathcal{I}_\ell(X_2; Y)\Big)$$

$$- \sum_{\ell \in \mathcal{A}'} P_\ell \mathcal{I}_\ell(X_2; Y|X_1)$$

$$\stackrel{(c)}{=} \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}\setminus\left\{\mathcal{A}'' \bigcup \{2\}\right\}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) + \sum_{\ell \in \mathcal{A}''} P_\ell \mathcal{I}_\ell(X_2; Y, X_1) - \sum_{\ell \in \mathcal{A}'} P_\ell \mathcal{I}_\ell(X_2; Y, X_1)$$

$$\stackrel{(d)}{=} \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}\setminus\left\{\mathcal{A}'' \bigcup \{2\}\right\}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) + \sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell \mathcal{I}_\ell(X_2; Y, X_1) - \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell \mathcal{I}_\ell(X_2; Y, X_1)$$

$$\stackrel{(e)}{=} \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}\setminus\left\{\mathcal{A}'' \bigcup \{2\}\right\}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) + \sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell \mathcal{I}_\ell(X_2; Y|X_1)$$

$$- \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell \Big(\mathcal{I}_\ell(X_2; Y) + \mathcal{I}_\ell(X_2; X_1|Y)\Big)$$

$$= \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}\setminus\left\{\mathcal{A}'' \bigcup \{2\}\right\}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) - \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) + \sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell \mathcal{I}_\ell(X_2; Y|X_1)$$

$$- \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell \mathcal{I}_\ell(X_2; Y)$$

$$\stackrel{(f)}{=} \mathcal{I}(X_1, X_2; Y) + \sum_{\ell \in \mathcal{S}\setminus\left\{\mathcal{A}' \bigcup \mathcal{A}'' \bigcup \{2\}\right\}} P_\ell \mathcal{I}_\ell(X_2; X_1|Y) + \sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell \mathcal{I}_\ell(X_2; Y|X_1) - \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell \mathcal{I}_\ell(X_2; Y)$$

where:

(a), (c) and (e) follow from the fact that X_1 is independent of X_2 given $(U, \widetilde{S}_1, \widetilde{S}_2)$;

(b) follows from the fact that $\mathcal{A}'' \subseteq \mathcal{S}\setminus\{2\}$;

(d) follows by eliminating common factors from the last two sums;

(f) follows from the fact that $\widetilde{\mathcal{A}}' \subseteq \mathcal{S}\setminus\left\{\mathcal{A}'' \bigcup \{2\}\right\}$, and by eliminating the common factors from the two relevant sums.

To conclude, applying the FME on (180) and (181) yields the following inequality:

$$\sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell R_{2\ell} + R_1' + R_2' - \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell R_{2\ell}$$

$$< \sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell \mathcal{I}_\ell(X_2;Y|X_1) + \mathcal{I}(X_1,X_2;Y) - \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell \mathcal{I}_\ell(X_2;Y) + \sum_{\ell \in \mathcal{S} \Big\backslash \{\mathcal{A}' \cup \mathcal{A}'' \cup \{2\}\}} P_\ell \mathcal{I}_\ell(X_2;X_1|Y) \tag{187}$$

Finally, note that by adding (175) taken for the subset $\mathcal{A}^{''}$ to (178) taken for the subset $\mathcal{A}^{'}$, we get:

$$\sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell R_{2\ell} + R_1' + R_2' - \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell R_{2\ell} < \sum_{\ell \in \widetilde{\mathcal{A}}''} P_\ell \mathcal{I}_\ell(X_2;Y|X_1) + \mathcal{I}(X_1,X_2;Y) - \sum_{\ell \in \widetilde{\mathcal{A}}'} P_\ell \mathcal{I}_\ell(X_2;Y)$$

which is tighter than (187) in view of the non-negativity of the mutual information. (187) is therefore also redundant.

We hence conclude that we are left with inequalities of the forms (175)-(178) for every $\mathcal{A} \subseteq \mathcal{S}_{1,2}$. This in turn implies that all the components which involve the element $\{2\}$ are eliminated. By repeating Stage 2 in order to eliminate $R_{23}, R_{24}, \ldots, R_{2k}$ (and their associated components), it can be shown that the only non-redundant inequalities are those taken for the subsets $\mathcal{A} \subseteq \mathcal{S}_{1,2,\ldots,k}$. However, since $\mathcal{S}_{1,2,\ldots,k} = \emptyset$, its only subset is $\mathcal{A} = \emptyset$, for which the inequalities in (175)-(178) reduce to:

$$R_2' < I(X_2;Y|X_1,U,S,\widetilde{S}_1 = \tilde{s}_1, \widetilde{S}_2), \tag{188}$$

$$R_1' + R_2' < I(X_1,X_2;Y|U,S,\widetilde{S}_1 = \tilde{s}_1, \widetilde{S}_2). \tag{189}$$

APPENDIX D
PROOF OF THE MARKOV RELATION IN (60)-(62)

We prove the Markov relation (60)-(62) using the following claims.

The equality in (60) follows from the facts that $(V_1^\ell, V_2^\ell, S^{q-d_1-1}) - S_{q-d_1} - S_{q-d_2} - S_q$ and $(V_1^\ell, V_2^\ell, S^{Q-d_1-1}) - (S_{Q-d_1}, Q) - S_{Q-d_2} - S_Q$.

To show (61) consider the following relations

$$\begin{aligned}
p(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, u_q, q) &= p(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) \\
&= \sum_{m_1 \in \mathcal{M}_1} p(m_1, x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) \\
&= \sum_{m_1 \in \mathcal{M}_1} p(m_1|s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, m_1, s^{q-d_1-1}, q) \\
&\qquad \cdot p(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, m_1, s^{q-d_1-1}, q) \\
&\overset{(a)}{=} \sum_{m_1 \in \mathcal{M}_1} p(m_1|s_{q-d_1} v_1^\ell, v_2^\ell, m_1, s^{q-d_1-1}, q) p(x_{1,q}|s_{q-d_1}, v_1^\ell, v_2^\ell, m_1, s^{q-d_1-1}, q) \\
&= \sum_{m_1 \in \mathcal{M}_1} p(m_1, x_{1,q}|s_{q-d_1}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) \\
&= p(x_{1,q}|s_{q-d_1}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q)
\end{aligned}$$

where (a) follows from the fact that M_1 is independent of S^n given (V_1^ℓ, V_2^ℓ) and the fact that $X_{1,q}$ is a deterministic function of $(V_1^\ell, V_2^\ell, M_1, S_{q-d_1}, S^{q-d_1-1})$. Now, since this is true for all q, and because the auxiliary RV is defined as $U = (V_1^\ell, V_2^\ell, S^{Q-d_1-1}, Q)$, it holds that

$$P(x_1|s, \tilde{s}_1, \tilde{s}_2, u) = P(x_1|\tilde{s}_1, u).$$

Finally, to show (62) we use the following relations

$$\begin{aligned}
p(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, u_q, q) &= p(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) \\
&= \sum_{m_2 \in \mathcal{M}_2} p(m_2, x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) \\
&= \sum_{m_2 \in \mathcal{M}_2} p(m_2|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) \\
&\qquad \cdot p(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, m_2, s^{q-d_1-1}, q) \\
&\overset{(a)}{=} \sum_{m_1 \in \mathcal{M}_1} p(m_2|s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) p(x_{2,q}|s_{i-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, m_2, s^{q-d_1-1}, q) \\
&= \sum_{m_1 \in \mathcal{M}_1} p(m_2, x_{2,q}|s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{i-d_1-1}, q) \\
&= p(x_{2,q}|s_{q-d_1}, s_{q-d_2}, v_1^\ell, v_2^\ell, s^{q-d_1-1}, q) \\
&= p(x_{2,q}|s_{q-d_1}, s_{q-d_2}, u_q)
\end{aligned}$$

where (a) follows from the fact that M_2 is independent of (X_1, S^n) given (V_1^ℓ, V_2^ℓ) and the fact that $X_{2,q}$ is independent of $(X_{1,q}, S_q)$ given $(V_1^\ell, V_2^\ell, M_2, S_{q-d_1}, S_{q-d_2}, S^{q-d_1-1})$. Again, since the above holds for every q and by the definition of U, we can conclude that

$$P(x_2|x_1, s, \tilde{s}_1, \tilde{s}_2, u) = P(x_2|\tilde{s}_1, \tilde{s}_2, u).$$

Appendix E
Proof of Lemma 6

In order to prove the RHS of (140), consider:

$$\begin{aligned}
\mathbb{P}\big\{(x^n,y^n)\in\mathcal{T}_\epsilon^{(n)}\big\} &= \sum_{(x^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} p(x^n,y^n)\\
&\overset{(a)}{=} \sum_{\substack{(x^{n_1},y^{n_1})\in\mathcal{T}_\epsilon^{(n_1)}\\(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}}} p(x^{n_1},y^{n_1},x^{n_2},y^{n_2})\\
&= \sum_{\substack{(x^{n_1},y^{n_1})\in\mathcal{T}_\epsilon^{(n_1)}\\(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}}} p(x^{n_1},y^{n_1})p(x^{n_2},y^{n_2}|x^{n_1},y^{n_1})\\
&\overset{(b)}{=} \sum_{(x^{n_1},y^{n_1})\in\mathcal{T}_\epsilon^{(n_1)}} p(x^{n_1},y^{n_1}) \sum_{(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}} p(x^{n_2})p(y^{n_2})\\
&\overset{(c)}{\le} 1\cdot \sum_{(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}} 2^{-n_2H(X)(1-\epsilon)}\cdot 2^{-n_2H(Y)(1-\epsilon)}\\
&\overset{(d)}{\le} 2^{n_2H(X,Y)(1+\epsilon)}\cdot 2^{-n_2H(X)(1-\epsilon)}\cdot 2^{-n_2H(Y)(1-\epsilon)}\\
&= 2^{n_2\left(H(X,Y)-H(X)-H(Y)+\epsilon\left(H(X,Y)+H(X)+H(Y)\right)\right)}\\
&\overset{(e)}{=} 2^{-n_2(I(X;Y)+\delta_\epsilon)}
\end{aligned}\tag{190}$$

where:

(a) follows from the fact that if $(x^n,y^n)\in\mathcal{T}_\epsilon^{(n)}$ then $(x^{n_1},y^{n_1})\in\mathcal{T}_\epsilon^{(n_1)}$ and $(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}$;

(b) follows from the fact that the sequences are drawn independently and the fact that the (x^{n_2},y^{n_2}) sequences are drawn according to $P_X(x)P_Y(y)$;

(c) and (d) follow from the fact that probability is always upper bounded by 1 and from the properties of typical sets;

(e) follows by denoting $\delta_\epsilon \triangleq \epsilon\big(H(X,Y)+H(X)+H(Y)\big)$.

We thus conclude that

$$\mathbb{P}\big\{(x^n,y^n)\in\mathcal{T}_\epsilon^{(n)}\big\} \le 2^{-n_2(I(X;Y)+\delta_\epsilon)}\tag{191}$$

where $\delta_\epsilon\to 0$ as $\epsilon\to 0$.

To prove the LHS of (140), we start from (b) in (190) and write:

$$\begin{aligned}
\mathbb{P}\big\{(x^n,y^n)\in\mathcal{T}_\epsilon^{(n)}\big\} &= \sum_{(x^{n_1},y^{n_1})\in\mathcal{T}_\epsilon^{(n_1)}} p(x^{n_1},y^{n_1}) \sum_{(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}} p(x^{n_2})p(y^{n_2})\\
&\overset{(a)}{\ge} (1-\delta) \sum_{(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}} p(x^{n_2})p(y^{n_2})
\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\geq} (1-\delta) \sum_{(x^{n_2},y^{n_2})\in\mathcal{T}_\epsilon^{(n_2)}} 2^{-n_2 H(X)(1+\epsilon)} \cdot 2^{-n_2 H(Y)(1+\epsilon)} \\
&\overset{(c)}{\geq} (1-\delta)\cdot(1-\delta_{\epsilon,n}) \cdot 2^{n_2 H(X,Y)(1-\epsilon)} \cdot 2^{-n_2 H(X)(1+\epsilon)} \cdot 2^{-n_2 H(Y)(1+\epsilon)} \\
&\overset{(d)}{=} (1-\delta'_{\epsilon,n}) \cdot 2^{n_2\left(H(X,Y)-H(X)-H(Y)-\epsilon\left(H(X,Y)+H(X)+H(Y)\right)\right)} \\
&\overset{(e)}{=} (1-\delta'_{\epsilon,n}) \cdot 2^{-n_2(I(X;Y)-\delta_\epsilon)}
\end{aligned}
$$

where:

(a) follows from the fact that (x^{n_1}, y^{n_1}) are drawn i.i.d according to $P_{XY}(x,y)$ and thus $Pr(x^{n_1}, y^{n_1} \in \mathcal{T}_\epsilon^{(n_1)}) \geq (1-\delta)$;

(b) and (c) follow from the properties of typical sets;

(d) follows from denoting $\delta'_{\epsilon,n} \triangleq \delta_{\epsilon,n}\delta - \delta_{\epsilon,n} - \delta$;

(e) follows by denoting $\delta_\epsilon \triangleq \epsilon\big(H(X,Y) + H(X) + H(Y)\big)$.

We conclude that

$$
\mathbb{P}\big\{(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}\big\} \geq (1-\delta'_{\epsilon,n}) \cdot 2^{-n_2(I(X;Y)-\delta_\epsilon)} \tag{192}
$$

where $\delta_\epsilon \to 0$ and $\delta'_{\epsilon,n} \to 0$ as $\epsilon \to 0$. The proof is completed by combining (191) and (192).

REFERENCES

[1] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice.* Wiley, 2009, vol. 3.

[2] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 441–445, May 1983.

[3] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inf. Theory*, Tsahkadsor, Armenia, U.S.S.R., Sept. 1973, pp. 23–52.

[4] H. Liao, "Multiple access channels," Ph.D. dissertation, Elec. Eng. Dept., Univ. Hawaii, Honolulu, 1972.

[5] D. Slepian and J. K. Wolf, "A coding theorem for multiple-access channel with correlated sources," *Bell Syst. Tech. J.*, vol. 51, pp. 1037–1076, 1973.

[6] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[7] F. M. J. Willems, "Information-theoretical results for the discrete memoryless multiple access channel," Ph.D. dissertation, KU Leuven, Leuven, Belgium, 1982, ph.D. Thesis.

[8] O. Simeone, D. Gündüz, H. V. Poor, A. J. Goldsmith, and S. Shamai (Shitz), "Compound multiple-access channels with partial cooperation," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2425–2441, June 2009.

[9] V. M. Prabhakaran and P. Viswanath, "Interference channels with destination cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 187–209, January 2011.

[10] ——, "Interference channels with source cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 156–186, January 2011.

[11] I.-H. Wang and D. Tse, "Interference mitigation through limited receiver cooperation: Symmetric case," in *IEEE Inf. Theory Workshop*, Taormina, Italy, October 2009, pp. 579–583.

[12] H. Bagheri, A. S. Motahari, and A. K. Khandani, "On the symmetric Gaussian interference channel with partial unidirectional cooperation," *IEEE Trans. on Wireless Communication*, 2009, submitted for publication.

[13] C. Ng, N. Jindal, A. J. Goldsmith, and U. Mitra, "Capacity gain from two-transmitter and two-receiver cooperation," *IEEE Trans Inf. Theory*, vol. 53, no. 10, pp. 3822–3827, April 2007.

[14] I. Maric, R. D. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans Inf. Theory*, vol. 53, no. 10, pp. 3536–3548, October 2007.

[15] I. Maric, A. J. Goldsmith, G. Kramer, and S. Shamai (Shitz), "On the capacity of interference channels with one cooperating transmitter," *European Transactions on Telecommunications*, vol. 19, pp. 405–420, 2008.

[16] R. Dabora and S. D. Servetto, "On the role of estimate-and-forward with time sharing in cooperative communication," *IEEE Trans Inf. Theory*, vol. 54, no. 10, pp. 4409–4431, October 2008.

[17] D. Gündüz and E. Erkip, "Source and channel coding for cooperative relaying," *IEEE Trans Inf. Theory*, vol. 53, no. 10, pp. 3454–3475, October 2007.

[18] L. Sankar, G. Kramer, and N. B. Mandayam, "Dedicated-relay vs. user cooperation in time-duplexed multiaccess networks," *Journal of Communications*, vol. 6, no. 4, pp. 330–339, July 2011.

[19] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Local base station cooperation via finite-capacity links for the uplink of linear cellular networks," *IEEE Trans Inf. Theory*, vol. 55, no. 1, pp. 190–204, January 2009.

[20] G. Kramer, I. Maric, and R. D. Yates, "Cooperative communications," *Foundations and Trends in Networking*, vol. 1, no. 3/4, 2006.

[21] A. Haghi, R. Khosravi-Farsani, M. R. Aref, and F. Marvasti, "The capacity region of p-transmitter/q-receiver multiple-access channels with common information," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7359–7376, November 2011.

[22] ——, "The capacity region of fading multiple access channels with cooperative encoders and partial CSIT," in *Proc. Int. Symp. Inf. Theory*, Austin, Texas., 2010.

[23] H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6379–6396, October 2011.

[24] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, no. 4, pp. 289–293, October 1958.

[25] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. and Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[26] C. Heegard and A. A. E. Gamal, “On the capacity of computer memory with defects,” *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 731–739, September 1983.

[27] A. J. Goldsmith and P. P. Varaiya, “Capacity of fading channels with channel side information,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, November 1997.

[28] G. Caire and S. Shamai (Shitz), “On the capacity of some channels with channel state information,” *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, September 1999.

[29] A. Lapidoth and Y. Steinberg, “The multiple-access channel with causal side information: Common state,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 32–50, January 2013.

[30] M. Li, O. Simeone, and A. Yener, “Leveraging strictly causal state information at the encoders for multiple access channels,” in *Proc. Int. Symp. Inf. Theory*, 2011.

[31] H. Viswanathan, “Capacity of Markov channels with receiver CSI and delayed feedback ,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, March 1999.

[32] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[33] H. S. Wang and N. Moayeri, “Finite-state Markov channel - a useful model for radio communication channels,” *IEEE Trans. Veh. Technol.*, vol. 44, p. 163171, February 1995.

[34] U. Basher, A. Shirazi, and H. H. Permuter, “Capacity region of finite state multiple-access channel with delayed state information at the transmitters,” *IEEE Trans Inf. Theory*, 2011, submitted for publication.

[35] H. H. Permuter and O. Simeone, “Source coding when the side information may be delayed,” *accepted to IEEE Trans. Inf. Theory*, 2011.

[36] V. Venkatesan, “Optimality of gaussian inputs for a multi-access achievable rate region,” Ph.D. dissertation, ETH Zurich, Switzerland, 2007.

[37] S. I. Bross, A. Lapidoth, and M. A. Wigger, “The Gaussian MAC with conferencing encoders,” in *Proc. Int. Symp. Inf. Theory*, July 2008, pp. 2702–2706.

[38] M. Wigger and G. Kramer, “Three-user MIMO MACs with cooperation,” in *IEEE Inf. Theory Workshop*, June 2009, pp. 221–225.

[39] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[40] T. Ando and D. Petz, “Gaussian Markov triplets approached by block matrices,” *Acta Math*, vol. 75, pp. 265–281, 2009.

[41] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programmingg, version 2.0 beta,” http://cvxr.com/cvx, September 2012.

[42] L. Nuaymi, *WiMAX: Technology for Broadband Wireless Access*. Wiley, 2007.

[43] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.